Statistical Models to Assess Associations between the Built Environment and Health: Examining Food Environment Contributions to the Childhood Obesity Epidemic

by

Jonggyu Baek

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2014

Doctoral Committee:

Associate Professor Brisa N. Sánchez, Chair
Assistant Professor Veronica J. Berrocal
Professor Timothy D. Johnson
Assistant Professor Sung Kyun Park

To Hyesung and my parents

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Brisa N. Sánchez for her continuous support of my Ph.D study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me all the time in researching and writing this thesis. I could not have imagined a better advisor and mentor for my Ph.D study. Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Veronica J. Berrocal, Dr. Timothy D. Johnson, and Dr. Sung Kyun Park, for their encouragement and insightful comments.

My sincere thanks also go to the Brain Attack Surveillance in Corpus Christi (BASIC) project team for offering me great opportunities in their research groups and letting me research diverse exciting projects, and Dr. Emma V. Sanchez-Vaznaugh for sharing her insightful knowledge and suggestions of childhood obesity research.

Last but not least, I would like to thank my family: my parents, SeungRak Baek and Jinsook Kim; my parents-in-law, ManSik Min and MoonHee Ko; and my wife, Hyesung Min, for their love and support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Statistical Models to Assess Associations between the Built Environment and Health: Examining Food Environment Contributions to the Childhood Obesity Epidemic

by

Jonggyu Baek

Chair: Brisa N. Sánchez

Models are developed and applied to examine the associations between built environment features and health. These developments are motivated by studies examining the contribution of features of the built food environment near schools, such as availability of fast food restaurants and convenience stores, to children's body weight. The data used in this dissertation come from a surveillance database that captures body weight and other characteristics for all children in 5^{th}, 7^{th}, and 9^{th} grades enrolled in public schools in California during 2001-2010 and a commercial data source that contains the locations of all food establishments in California for the same time period. First, we develop a hierarchical multiple informants model (HMIM) for clustered data that estimates the marginal association of multiple built environment features and formally tests if the strength of their association differs with the outcome. Using this new model, we establish that the contribution of the availability of convenience stores to children's body mass index z-scores (BMIz) is stronger than that of fast

food restaurants. Second, we propose to use a distributed lag model (DLM) to examine whether and how the association between the number of convenience stores and children's BMIz decays with longer distance from schools. In this model, distributed lag (DL) covariates are the number of convenience stores within several contiguous "ring"-shaped areas from schools rather than circular buffers, and their coefficients are modeled as a function of distance, using smoothing splines. We find that associations are stronger with closer proximity to schools and vanish by about 2 miles from school locations. Third, we develop a hierarchical distributed lag model (HDLM) to systematically examine the variability of the built environment association across regions to help address a yet unanswered question in the built environment literature: whether and how activity spaces relevant to health vary across regions. We find DL coefficients vary across regions, implying that variation in activity spaces also exists. We also identify areas where children's BMIz is more vulnerable to built environment factors. This dissertation provides novel methods with which to study how built environment factors affect health.

CHAPTER I

Introduction

Neighborhood resources and built environment factors have recently received great attention as potential contributors to health beyond individual factors because humans and the built environment surrounding them are interrelated. For instance, humans may change the built environment to suit their needs and at the same time the built environment directly constrains individuals' behaviors and their choices (*Cummins et al.*, 2007; *Diez-Roux*, 1998; *Susser*, 1994). Food availability and availability of parks or recreational facilities constrain food choices and ability to exercise. However, the association between multiple environmental factors and individual health outcomes can be difficult to untangle. Recently, advancements of geographic information systems (GIS) have enabled researchers to more scientifically examine how features of the built environment are associated with health outcomes, such as childhood obesity (*Pearce et al.*, 2006). By integrating spatial information from a range of disparate sources into a single database and developing precise measures of the built environment, *Thornton et al.* (2011) illustrated the usefulness of GIS for built environment research.

Given the now widely known childhood obesity epidemic, one related research question is whether and how features of the built environment within children's time-activity spaces impact childhood obesity. Children spend a large proportion of their

time in and around schools and therefore commercial establishments near schools offering "junk" foods have been considered as possible contributors to the childhood obesity epidemic. The underlying idea is that availability of establishments that sell high energy, low nutrition foods near schools may increase consumption of junk foods, both through direct availability and purchasing on the way to and from schools, as well as indirectly through exposure to advertising that may shape individuals' choices (*Gebauer and Laska*, 2011; *Hillier et al.*, 2009).

These features of the food environment are typically measured by the number of food stores within a specific distance from schools (e.g., number of stores falling within a circle of 1 mile radius around schools, also known as a 1 mile "buffer") or a distance of 15 minutes by walk. However, not all studies show consistent results (*Alviola et al.*, 2014; *Currie et al.*, 2009; *Harris et al.*, 2011; *Langellier*, 2012; *Sánchez et al.*, 2012) regarding the association between number of food stores around schools and children's weight status. Part of the reasons for this inconsistency may be due to differences in how the measures of the built environment are constructed, as noted by many authors (*Apparicio et al.*, 2008; *Flowerdew et al.*, 2008); for instance, the buffer size within which the environment features are counted. The inconsistency of the results to the choice of geographic scale is often referred to as the "Modifiable Areal Unit Problem" (MAUP) (*Openshaw*, 1984; *Fotheringham and Wong*, 1991; *Openshaw*, 1996). Thus, it is important to determine the most appropriate buffer size.

To determine the buffer size that most strongly influences children's health, previous studies have compared associations among a-priori chosen buffer sizes by examining the overlap of the corresponding confidence intervals, or alternatively, by examining the distance at which the associations are, or cease to be, significant (*Davis and Carpenter*, 2009). Additionally, studies have compared the goodness-of-fits from several fitted models (*Guo and Bhat*, 2004; *Leal et al.*, 2011; *Vallée and Chauvin*, 2012). However, comparing associations from separately fitted models based on confidence

intervals may yield misleading results since estimated associations of measured factors on the same outcome are treated as being independent, although they are in fact correlated. Moreover, *Spielman and Yoo* (2009) showed through a simulation that comparing several models, each using different buffer sizes, based on goodness-of-fit statistics may not be useful in identifying the best underlying buffer size.

Furthermore, the true underlying buffer size may depend on various factors, such as types of food stores, individual characteristics (including age, race/ethnicity, and gender), area characteristics, and time at which the data are collected. From a policy perspective, comparing the strength of associations between two competitive types of food stores on children's health may be of interest. Regulating the density of "junk" food stores near schools, for instance, has been suggested as a possible intervention (*Austin et al.*, 2005). Investigating how associations of the built environment on children's health differ by participant characteristics, such as age, can also help understand underlying mechanisms of how children interact with features of the built environment, hence, improving our understanding of who is more vulnerable to these exposures and developing tailored interventions. At the areal level, recent empirical studies have found, for example, that the perceived neighborhood size is smaller for residents of peripheral city neighborhoods compared to downtown neighborhoods (*Vallée et al.*, 2014). When neighborhood level data are not available, it may still be possible to indirectly assess the variation on the relevant distances for a given health outcome.

Motivated by the FitnessGram data set examining the association between multiple features of the built environment, namely the presence of different types of food stores near schools, and children's weight, the first part of this dissertation extends the multiple informant model (MIM) developed independently by *Pepe et al.* (1999) and *Horton et al.* (1999), to account for hierarchical data. Because children are nested within schools, hierarchical multiple informant models (HMIMs) enable formal tests

of marginal associations among multiple features of the environment (e.g. fast food restaurants and convenience stores) and the same outcome (child's health). This type of testing procedure is amenable when comparing marginal associations from each separate regression model on the same outcome, and when comparing the conditional association of one informant with an outcome adjusting for the other informant is not meaningful. The HMIM provides valid inference of testing associations estimated from separate hierarchical models on the same outcome while increasing efficiency by accounting for a hierarchical data structure. This HMIM is also appealing because it can be fully implemented in available statistical software for generalized estimating equation (GEE) methods. The HMIM is applied to formally test how the association between a child's BMI z-score (BMIz) and the number of food stores around schools changes over three pre-specified distances widely used in children's built environment studies, and how marginal associations of convenience stores are different from those of fast food restaurants on the child's outcome.

The second part of this thesis proposes to use distributed lag models (DLMs) to examine how the association between features of the built environment (e.g., the number of convenience stores) and health (e.g., children's BMIz) varies over distance from locations of interest (e.g., schools). Additionally, these models can be used to quantify the associations between health and features of the built environment within a-priori specified distances (e.g., 1/4, 1/2, or 1 miles) from study locations, to enable comparisons with existing approaches. The proposed model improves upon traditional linear models employed in epidemiology, ecology, and transport geography literature. These traditional models assume constant effects up to a-priori specified distances and assume that no further association exists beyond those distances. Fitting DLMs enables us to estimate the association between the measured feature and the outcome more accurately than using the traditional models; moreover, the assumptions of DLMs are less restrictive than those of the currently used models. DLMs are applied

to examine the effects of the food environment around schools on children's BMIz using FitnessGram data and to investigate how the built environment association differs by individual characteristics (e.g., 5^{th} graders vs. 7^{th} graders) possibly due to different health behaviors as children get older. Although DLMs have a long history in econometrics and other literature, this dissertation presents the first application of DLMs to study the associations between health and the built environment.

The third part of this thesis proposes a hierarchical distributed lag models (HDLM) to investigate whether there is variability in the association of the built environment with health across regions. It has long been argued in the built environment literature that the associations between the built environment and health outcomes might have different distance lags in different regions and are likely patterned across regions; however, to date, such variation has not been systematically examined in this context. Since variation in the DL coefficients can be partially attributed to larger or shorter relevant distances at given regions, the proposed HDLM can help shed light on this built environment conundrum using an agnostic approach. The HDLMs are implemented to analyze data on children's BMIz collected over 10 years from all California 7^{th} graders and examine variations in the distributed lag coefficients (i.e., the association between features of the built environment and children's BMIz) across regions, while accounting for correlation within schools. Whereas the HDLMs available in the literature have typically been estimated using two-stage approaches, we jointly estimate the DL coefficients across regions of interest to borrow strength across regions and stabilize variance of the coefficients. The estimation is carried out within a Bayesian framework using a shrinkage slice sampling technique because of its advantages over Metropolis-Hasting algorithms, e.g., the convergence rate of posterior samples and no need for controlling for acceptance rates in posterior samples.

The dissertation is structured as follows. In Chapter 2, we present a review of the MIM and provide all details concerning the development of the HMIM. Chapter 3 is

devoted to the application of DLM to the built environment research. In Chapter 4, the DLM is further extended to examine variability of DL coefficients across regions. Conclusions and future work are outlined in Chapter 5. The Appendix lists some technical details and the figures of the three chapters.

CHAPTER II

Hierarchical multiple informants models: Examining food environment contributions to the childhood obesity epidemic

2.1 Introduction

The childhood obesity epidemic has led several researchers to examine factors beyond the individual as possible causes of obesity. For instance, since children spend large amounts of time in schools, there is increased interest in environmental factors in or around schools. The presence of food stores such as fast food restaurants or convenience stores has received attention as children may purchase, or be exposed to advertising of energy-dense, nutrient-poor foods on their way to or from school. These features of the environment near schools are typically operationalized as the number of food stores within a specific distance from schools (e.g., number of stores falling within a circle of 1/2 mile radius around schools, also known as 1/2 mile "buffer"). The associations between each type of built environment features and children's body weight are examined in separate models because the marginal association between each feature of the built environment and children's body weight is of substantive interest or because features are strongly correlated making it difficult to include them simultaneously in one model. Comparing the strength of associations between one

food store type and another is of particular interest (e.g., convenience stores vs. fast food restaurant) because limiting certain types of food stores vs. others may need to be considered from a policy perspective. An overarching limitation of the methods presently employed in these prior studies is that they do not rigorously examine or test differences among associations between environment features and outcomes. Furthermore, the specific distance from schools at which the circle is drawn (also known as buffer size) is typically chosen in an ad hoc manner, since the distance from schools at which the presence of food stores may influence children's body weight is unknown. Some researchers have compared associations of the outcome with a feature at different buffer sizes by examining the extent of overlap of confidence intervals for the associations obtained from several buffer sizes or examined the distance at which the association is, or ceases to be, significant (*Davis and Carpenter*, 2009). However, comparing the extent of overlap of confidence intervals is problematic because the estimates are correlated. The purpose of the present research is to develop a hierarchical multiple informant model (HMIM) that will facilitate comparing differences in the associations of the same type of food store at several buffers sizes from schools, and/or differences in the associations of two or more types of food stores.

Methods for multiple informant data were independently proposed by *Horton et al.* (1999) and *Pepe et al.* (1999) and have been comprehensively reviewed by *Horton and Fitzmaurice* (2004). The term "multiple informants" refers to information from multiple sources used to measure the same construct. *Horton et al.* (1999) give an example of multiple informants, such as information collected from a child's teacher and parent to assess the child's psychopathology. In our setting, the multiple informant predictors are features of the environment (e.g., multiple store types or number of a given store type at several buffer sizes) that may affect children's weight.

Models for multiple informants can be constructed using non-standard generalized estimating equation (GEE) methods to estimate the marginal association between

each multiple source predictor and an outcome, and provide a formal comparison of the strength of the associations between each predictor with the outcome (*Horton et al.*, 1999; *Pepe et al.*, 1999). Alternatively, *Litman et al.* (2008) developed a maximum likelihood estimation (MLE) approach that, under a joint normality of predictors and an outcome, can accommodate more general models than can be estimated with a GEE method. The MLE approach can incorporate multiple informants measured in different scales and enable estimation of a common "standardized" association (e.g., adjusted correlation coefficient), and incorporate data missing at random. However, existing multiple informant methods are limited to non-hierarchical data where univariate outcomes are measured on independent subjects. Although *Horton and Fitzmaurice* (2004) stressed the importance of complex survey designs, the estimating equations they employed assume independent subjects.

In Section 2.2 we briefly review multiple informant methods for univariate outcomes, and extend multiple informant approaches to a hierarchical data setting in Section 2.3. In Section 2.4 we present a small scale simulation study to highlight properties of the proposed methods. In Section 2.5 we apply the methods to examine the association between the presence of food stores near schools and child's body mass index z-score (BMIz) using a surveillance dataset from all $5^{th}, 7^{th}$, and 9^{th} grade children enrolled in public schools in the State of California. We use two different features of the food environment: fast food restaurants and convenience stores. Section 2.6 concludes with a discussion.

2.2 Review of univariate multiple informant models and GEEs

2.2.1 Non-standard GEE approach for Multiple Informant Models with Independent Subjects

Based on a non-standard application of GEE methods, *Pepe et al.* (1999) and *Horton et al.* (1999) developed a multiple informant model (MIM) to estimate the association between univariate outcomes and multiple informant predictors. For the $i^{th}(i = 1, \ldots, n)$ subject, let Y_i be an outcome and X_{ik} be multiple informants, $k = 1, \ldots, K$. The marginal associations between the outcome and each predictor, X_{ik}, are defined by separate regressions

$$E[Y_i|X_{ik}] = \beta_{0k} + \beta_{1k}X_{ik}, \tag{2.1}$$

where β_{0k} and β_{1k} are the intercept and slope parameters in the k^{th} regression, $k = 1, \ldots, K$. Joint estimation of model parameters can be accomplished by re-structuring the data as

$$\tilde{\mathbf{Y}}_i = \begin{bmatrix} Y_i \\ Y_i \\ \vdots \\ Y_i \end{bmatrix}, \ \tilde{\mathbf{X}}_i = \begin{bmatrix} 1 & X_{i1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & X_{i2} & \cdots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & X_{iK} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \\ \vdots \\ \beta_{0K} \\ \beta_{1K} \end{bmatrix}. \tag{2.2}$$

Note that $\tilde{\mathbf{Y}}_i$ has K copies of the same outcome Y_i, and covariate vectors, $[1\ X_{ik}], k = 1, \ldots, K$, are diagonally stacked in $\tilde{\mathbf{X}}_i$; correspondingly, $\boldsymbol{\beta}$ is a vector with all coefficients β_{0k} and β_{1k} stacked. Essentially, each subject is treated as an independent

cluster with K repeated measures (which are in fact K copies of the same outcome). Under the assumption of the identity link, constant variance and the working independence correlation matrix, the GEE for $\boldsymbol{\beta}$ is

$$\Sigma_{i=1}^{n} \tilde{\mathbf{X}}_i^T (\tilde{\mathbf{Y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta}) = 0. \tag{2.3}$$

By solving (2.3), the regression parameters $\boldsymbol{\beta}$ can be estimated, and the variance-covariance matrix for the $2K$ parameter estimates, $\hat{\boldsymbol{\beta}}$, can be derived by either the empirical variance estimator or the model-based variance via the GEE approach (*Litman et al.*, 2008). Since the multiple informant model basically employs GEE with re-structured data, binary or count data can be also fitted by changing the link function (e.g., logit, log) (*Liang and Zeger*, 1986).

Litman et al. (2008) demonstrated that assuming the working independence correlation is optimal for certain models since the non-standard GEE approach and MLE approach yield the same estimator. Further, the working independence structure within cluster is necessary to ensure consistency in the non-standard GEE approach (*Pan et al.*, 2000; *Pepe and Anderson*, 1994). Indeed, without a zero constraint for off-diagonal terms, joint modeling of the same outcome on multiple informants is invalid. For instance, suppose that we have an outcome y_{it} and a predictor x_{it} for the i^{th} subject at two occasions, $t = 1, 2$. Under a normal assumption of y_{it} conditional on x_{it}, the joint distribution of y_{it} given x_{it} for $t = 1, 2$ can be expressed as

$$\begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix} \Big| x_{i1}, x_{i2} \sim N\left(\begin{bmatrix} \beta_{01} + \beta_{11} x_{i1} \\ \beta_{02} + \beta_{12} x_{i2} \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right), \tag{2.4}$$

An implicit assumption of GEE is that covariates at a given occasion are not related to the outcome given the same covariate measured at another occasion, i.e., $E[y_{i1}|x_{i1}, x_{i2}] = E[y_{i1}|x_{i1}]$ and $E[y_{i2}|x_{i1}, x_{i2}] = E[y_{i2}|x_{i2}]$ (*Pan et al.*, 2000; *Pepe and*

Anderson, 1994). Another implicit assumption here is $y_{i1} \neq y_{i2}$. When $y_{i1} = y_{i2}$, the non-standard GEE approach needs to impose a zero constraint on σ_{12} (all off-diagonal terms in the covariance matrix).

In our motivating study, we are interested in the associations between multiple correlated predictors and weight status of children nested in schools (i.e., hierarchical data). We next review hierarchical modeling using well-developed GEE methods and subsequently extend the multiple informant model to hierarchical data.

2.2.2 The GEE Model with Exchangeable Correlation Structure

GEE methods have been well established and are extensively used to model hierarchical data. We briefly review the specific case of GEE with an exchangeable correlation structure as a building block for our proposed models in Section 2.3. Consider a simple case where data consist of J clusters, each with n_j units with measures on an outcome and a covariate: $\{y_{ij}, x_{ij}\}, i = 1, \ldots, n_j$ for each of $j = 1, 2, \ldots, J$ clusters. Units are assumed to be correlated within clusters, but independent across clusters. A common correlation structure used for this data is an exchangeable correlation−i.e., $corr(y_{ij}, y_{i'j}) = \rho, i \neq i'$ in the j^{th} cluster. A generalized linear model is commonly used to relate the mean of $y_{ij}, \mu_{ij} = E[y_{ij}]$, to a covariate, x_{ij}, via a link function $g(\cdot)$

$$g(u_{ij}) = \beta_0 + \beta_1 x_{ij}. \tag{2.5}$$

and the variance of y_{ij} is $Var(y_{ij}) = \phi v(\mu_{ij})$, where $v(\cdot)$ is a known variance function and ϕ is a dispersion parameter. Similar to (2.3), GEE estimates, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)^T$, are given by solving

$$\Sigma_{j=1}^{J} \mathbf{D}_j \mathbf{V}_j^{-1} (\mathbf{Y}_j - \boldsymbol{\mu}_j) = 0. \tag{2.6}$$

where $\mathbf{Y}_j = (y_{1j}, \ldots, y_{n_j,j})^T$, $\boldsymbol{\mu}_j = E[\mathbf{Y}_j]$, $\mathbf{D}_j = \partial \boldsymbol{\mu}_j / \partial (\beta_0, \beta_1)^T$, $\mathbf{V}_j = \mathbf{A}_j^{1/2} \mathbf{R}_j \mathbf{A}_j^{1/2}$, $\mathbf{A}_j = \phi diag\{v(\boldsymbol{\mu}_{1j}), \ldots, v(\boldsymbol{\mu}_{n_j,j})\}$, and $\mathbf{R}_j$ is a working correlation matrix (*Liang and*

Zeger, 1986). For a continuous outcome y_{ij} with the identity link function and an exchangeable correlation assumption the solution of (2.6) for β with a known ρ is

$$\hat{\boldsymbol{\beta}} = (\Sigma_{j=1}^{J} \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j)^{-1} (\Sigma_{j=1}^{J} \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{Y}_j). \tag{2.7}$$

The empirical or 'sandwich' variance of $\hat{\boldsymbol{\beta}}$ is

$$\hat{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{BFB} \tag{2.8}$$

where $\mathbf{B} = (\Sigma_{j=1}^{J} \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j)^{-1}$, $\mathbf{F} = \Sigma_{j=1}^{J} \mathbf{X}_j^T \mathbf{V}_j^{-1} (\mathbf{Y}_j - \hat{\boldsymbol{\mu}}_j)(\mathbf{Y}_j - \hat{\boldsymbol{\mu}}_j)^T \mathbf{V}_j^{-1} \mathbf{X}_j$.

Since ϕ and ρ are generally unknown, $\hat{\boldsymbol{\beta}}$ needs to be iteratively re-estimated to update the estimated variance-covariance matrix, $\mathbf{V}_j$. To estimate the dispersion parameter ϕ and correlation ρ, refer to *Liang and Zeger* (1986). Since the empirical variance estimator (2.8) protects against a misspecified working correlation and variance structure, inference based on GEE estimators is robust to departures from the true covariance structure (*Litman et al.*, 2008; *Liang and Zeger*, 1986).

In Section 2.2.1, it is necessary to assume the working independence structure for a MIM for consistency of the estimators, but for hierarchical models reviewed here, the working independence assumption may be inefficient in some situations. *Mancl and Leroux* (1996) demonstrated that loss of efficiency for the working independence correlation assumption can be substantial even for small correlation when the coefficient of variation in the cluster sizes (CV) is greater than 0.5. In our motivating data, the number of children varies largely across schools (CV $\approx$ 1.1). We extend the MIM to hierarchical data structures by incorporating a block diagonal working correlation to make the model valid, but with diagonal blocks of exchangeable correlation structures to model correlations within clusters to enhance efficiency.

2.3 Hierarchical Multiple Informant Model (HMIM)

2.3.1 Data Structure and Model

Let Y_{ij} be an outcome of the i^{th} unit (e.g., child's BMIz) within the j^{th} cluster (e.g., school), and denote the mean of Y_{ij} as $\mu_{ij} = E[Y_{ij}]$, $i = 1, \ldots, n_j$ for each of $j = 1, 2, \ldots, J$. For simplicity, assume there are two multiple informant predictors measured at the cluster level, X_{j1} and X_{j2} (e.g., X_{j1} is the number of FFR and X_{j2} is the number of convenience store (CS) within d miles from the j^{th} school). Given a link function $g(\cdot)$, μ_{ij} can be modeled as

$$\begin{cases} g(\mu_{ij}) = \beta_{01} + \beta_{11} X_{j1} \\ g(\mu_{ij}) = \beta_{02} + \beta_{12} X_{j2} \end{cases}, \tag{2.9}$$

where β_{0k} and β_{1k} for $k = 1, 2$ are the population-level intercept and slope parameters for the k^{th} regression.

Similar to (2), the data are re-structured as

$$\tilde{\mathbf{Y}}_j = \begin{bmatrix} Y_{1j} \\ \vdots \\ Y_{n_j,j} \\ Y_{1j} \\ \vdots \\ Y_{n_j,j} \end{bmatrix}, \ \tilde{\mathbf{X}}_j = \begin{bmatrix} 1 & X_{j1} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{j1} & 0 & 0 \\ 0 & 0 & 1 & X_{j2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & X_{j2} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^1 \\ \boldsymbol{\beta}^2 \end{bmatrix}. \tag{2.10}$$

Note that two copies of the outcome vector $\mathbf{Y}_j = (Y_{1j}, \ldots, Y_{n_j,j})^T$ for all subjects $i = 1, 2, \ldots, n_j$ within the j^{th} cluster are stacked. The covariate matrices $\mathbf{X}_{jk}, k = 1, 2$, consist of n_j copies of the vector $[1 \ X_{jk}]$, and are diagonally stacked in $\tilde{\mathbf{X}}_j$. Accordingly, $\boldsymbol{\beta}$ contains all β_{0k}, β_{1k}, the population-level intercept and slope parameters.

Including individual level predictors and other cluster level variables is straight-

forward. For instance, let $\mathbf{W}_{ij} = [\mathbf{Z}_{ij}\ \mathbf{Z}_j]$ include individual level predictors $\mathbf{Z}_{ij}$ and other cluster level variables $\mathbf{Z}_j$. Then, the covariate matrices, $[\mathbf{X}_{j1}\ \mathbf{W}_{ij}]$ and $[\mathbf{X}_{j2}\ \mathbf{W}_{ij}]$, within the j^{th} cluster can be re-structured as in (2.10).

Careful modeling of the correlation within clusters can improve inference of the population level parameters $\boldsymbol{\beta}$. However, due to the implicit assumptions of the GEE as discussed in Section 2.2.1, we restrict the working covariance structure for the HMIM to a block diagonal matrix where the diagonal blocks are the correlation structures given each correlated predictor. Let $\mathbf{V}_{jk} = \phi_k \mathbf{R}_{jk}$, where $\mathbf{V}_{jk}$ consists of constant variance and a correlation matrix $\mathbf{R}_{jk}$ for $k = 1, 2$. In the motivating example, we use an exchangeable correlation structure with correlation ρ_k to model $\mathbf{R}_{jk}$, since children within schools can be assumed exchangeable. Hence, the working covariance matrix of the HMIM can be

$$\tilde{\mathbf{V}}_j = \begin{bmatrix} \mathbf{V}_{j1} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{j2} \end{bmatrix}_{(2n_j \times 2n_j)}. \tag{2.11}$$

Note that $\tilde{\mathbf{V}}_j$ consists of a block diagonal of distinct exchangeable covariance matrices, $\mathbf{V}_{j1}$ and $\mathbf{V}_{j2}$, given X_{j1} and X_{j2}, respectively.

With $\tilde{\mathbf{V}}_j$ and the identity link function, and by virtue of the block diagonal covariates and covariance matrices, the estimator for $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\Sigma_{j=1}^{J} \tilde{\mathbf{X}}_j^T \tilde{\mathbf{V}}_j^{-1} \tilde{\mathbf{X}}_j)^{-1} (\Sigma_{j=1}^{J} \tilde{\mathbf{X}}_j^T \tilde{\mathbf{V}}_j^{-1} \tilde{\mathbf{Y}}_j).$$

yields equivalent estimates to fitting a separate model for each predictor, $\hat{\boldsymbol{\beta}}^1, \hat{\boldsymbol{\beta}}^2$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^1 \\ \hat{\boldsymbol{\beta}}^2 \end{bmatrix} = \begin{bmatrix} (\Sigma_{j=1}^{J} \mathbf{X}_{j1}^T \mathbf{V}_{j1}^{-1} \mathbf{X}_{j1})^{-1} (\Sigma_{j=1}^{J} \mathbf{X}_{j1}^T \mathbf{V}_{j1}^{-1} \mathbf{Y}_j) \\ (\Sigma_{j=1}^{J} \mathbf{X}_{j2}^T \mathbf{V}_{j2}^{-1} \mathbf{X}_{j2})^{-1} (\Sigma_{j=1}^{J} \mathbf{X}_{j2}^T \mathbf{V}_{j2}^{-1} \mathbf{Y}_j) \end{bmatrix} \tag{2.12}$$

The empirical or 'sandwich' variance-covariance for the estimated parameters is

$$\hat{Var}(\hat{\boldsymbol{\beta}}) = \tilde{\mathbf{B}}\tilde{\mathbf{F}}\tilde{\mathbf{B}}$$

where $\tilde{\mathbf{B}} = (\Sigma_{j=1}^{J}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{V}}_j^{-1}\tilde{\mathbf{X}}_j)^{-1}$, and $\tilde{\mathbf{F}} = \Sigma_{j=1}^{J}\tilde{\mathbf{X}}_j^T\tilde{\mathbf{V}}_j^{-1}(\tilde{\mathbf{Y}}_j - \tilde{\mathbf{X}}_j\hat{\boldsymbol{\beta}})(\tilde{\mathbf{Y}}_j - \tilde{\mathbf{X}}_j\hat{\boldsymbol{\beta}})^T\tilde{\mathbf{V}}_j^{-1}\tilde{\mathbf{X}}_j$.

Equivalently, if the models for each multiple informant are fitted separately,

$$\hat{Var}(\hat{\boldsymbol{\beta}}) = \tilde{\mathbf{B}}^*\tilde{\mathbf{F}}^*\tilde{\mathbf{B}}^* \tag{2.13}$$

where $\tilde{\mathbf{B}}^* = \begin{bmatrix} (\Sigma_{j=1}^{J}\mathbf{X}_{j1}^T\mathbf{V}_{j1}^{-1}\mathbf{X}_{j1})^{-1} & \mathbf{0} \\ \mathbf{0} & (\Sigma_{j=1}^{J}\mathbf{X}_{j2}^T\mathbf{V}_{j2}^{-1}\mathbf{X}_{j2})^{-1} \end{bmatrix}$ and $\tilde{\mathbf{F}}^* = \sum_{j=1}^{J}\begin{bmatrix} \mathbf{X}_{j1}^T\mathbf{V}_{j1}^{-1}\mathbf{r}_j^1 \\ \mathbf{X}_{j2}^T\mathbf{V}_{j2}^{-1}\mathbf{r}_j^2 \end{bmatrix}\begin{bmatrix} \mathbf{X}_{j1}^T\mathbf{V}_{j1}^{-1}\mathbf{r}_j^1 \\ \mathbf{X}_{j2}^T\mathbf{V}_{j2}^{-1}\mathbf{r}_j^2 \end{bmatrix}^T$, where $\begin{bmatrix} \mathbf{r}_j^1 \\ \mathbf{r}_j^2 \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_j - \mathbf{X}_{j1}\hat{\boldsymbol{\beta}}^1 \\ \mathbf{Y}_j - \mathbf{X}_{j2}\hat{\boldsymbol{\beta}}^2 \end{bmatrix}$.That is, the empirical variance/covariance for $\hat{\boldsymbol{\beta}}$ can be calculated using results from each fitted marginal GEE model. From a practical point of view, fitting each marginal GEE model has computational efficiencies: 1) the dimension of the data will be smaller for any one model and 2) available GEE software can be implemented to obtain the empirical covariance matrix. Example R code for calculating the empirical variance/covariance matrix (2.13) is provided in Appendix A.

2.3.2 Hypothesis Testing

One advantage of the HMIM is that it gives a formal test to compare the association among multiple predictors on a univariate outcome while taking into account the correlation within clusters. In our motivating example, we seek to compare the association of two different features of the food environment (fast food restaurants (FFR) vs. convenience stores (CS)) with BMIz and to compare the associations between the number of FFR (or CS) and child's BMIz across several buffers. These tests can be conducted using general linear hypotheses expressed in the form

$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{L}_0$, where $\mathbf{L}$ consists of l linearly independent constraints on $\boldsymbol{\beta}$ and $\mathbf{L}_0$ is a vector of constant terms (usually a zero vector). The Wald test statistic is $T = (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}_0)^T(\mathbf{L}Var(\hat{\boldsymbol{\beta}})\mathbf{L}^T)^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}_0)$ which, given the asymptotic normality of $\hat{\boldsymbol{\beta}}$, asymptotically follows a chi-squared distribution with l degrees of freedom. We next describe a strategy to conduct hypothesis tests.

Two approaches can be followed to compare the associations between two (or more) predictors on the outcome (e.g., different features of food environment, $X_1 =$ fast food restaurants and $X_2 =$ convenience stores on child's BMIz within a given buffer size). The first is to use the predictors in their original scales and test for equality of coefficients, $H_0 : \beta_{11} = \beta_{12}$. Alternatively, if the scales are different (e.g., there is an overall preponderance of one feature compared to the other), the predictors can be standardized so that the coefficients are in standard deviation units (i.e., one standard deviation increase, or interquartile range increase). If we fail to reject the null hypothesis that the effects of multiple informants are the same, then, as suggested by *Litman et al.* (2008), a constrained model (i.e., a model that assumes $\beta_{11} = \beta_{12}$) could be used to increase power.

In our motivating example, we are also interested in comparing the effects of a given environmental feature (e.g., FFR) across several buffers on child's BMIz. Suppose that there are a priori specified distances of interest, $r_1 < r_2 < \cdots < r_K$, from schools, and let $X_1, X_2, \ldots, X_K$ be number of FFR within the corresponding buffers. For exposition suppose $K = 3$. Then, let $\beta_{1k}, k = 1, 2, 3$, be the corresponding marginal regression coefficients. We are interested in testing whether the effects differ, i.e., the overall test $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ vs. $H_1 :$ at least one differs. Failure to reject the null hypothesis suggests that the most appropriate buffer size is at least up to r_3 miles from schools. However, if the overall null hypothesis is rejected, we suggest the following subsequent tests. First, test the one-sided null hypothesis, $H_0 : \beta_{11} \leq \beta_{12}$. If the null hypothesis is rejected, then we decide that the buffer size r_1 miles from

schools has the strongest association, and testing stops. Otherwise, conduct a second one-sided test $H_0 : \beta_{12} \leq \beta_{13}$. If the null hypothesis is rejected, stop and conclude the buffer with size r_2 has strongest effects. Otherwise, buffer with size r_3 is most relevant.

2.3.3 AR(1) and Three-level Nested Structures

Given our motivating example we consider hierarchical structures where individuals are nested in larger units in which an exchangeable correlation structure is natural. However, hierarchical data can also arise in longitudinal repeated measures for which other correlation structures may be better suited. The HMIM can be applicable to this setting as well. For instance, let Y_{it} be the i^{th} child's BMIz at time t, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, and the two covariates be X_{it1} and X_{it2}. Data can be re-structured in a similar manner as (2.10)

$$\tilde{\mathbf{Y}}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{i,T} \\ Y_{i1} \\ \vdots \\ Y_{i,T} \end{bmatrix}, \ \tilde{\mathbf{X}}_i = \begin{bmatrix} 1 & X_{i11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{iT1} & 0 & 0 \\ 0 & 0 & 1 & X_{i12} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & X_{iT2} \end{bmatrix}.$$

Note that $\tilde{\mathbf{Y}}_i$ has two copies of the vector of repeated measures $\mathbf{Y}_i^T = [Y_{i1} \ \cdots \ Y_{iT}]$ and $\tilde{\mathbf{X}}_i$ has a block diagonal structure of the two covariates including the intercepts. To reflect within-cluster correlation over time in a longitudinal study, an AR(1) correlation structure can be used

$$\tilde{\mathbf{V}}_i = \begin{bmatrix} \sigma_1^2 AR(1, \rho_1)_{T \times T} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 AR(1, \rho_2)_{T \times T} \end{bmatrix}.$$

where $AR(1,\rho_1)$ and $AR(1,\rho_2)$ are $AR(1)$ correlation structures with an autocorrelation ρ_1 and ρ_2, respectively, and the variances are σ_1^2 and σ_2^2. Other variances can also be incorporated (e.g., a non-constant variance if an outcome is binary).

Another extension is for a three-level nested structure. For instance, suppose Y_{ijl} is the i^{th} child's BMI in the j^{th} school in the l^{th} county, $i = 1,\ldots,n_{jl}, j = 1,\ldots,n_l, l = 1,\ldots,L$. Assume that there are two county-level covariates $X_{lk}, k = 1,2$, of interest for comparison. With the independence assumption across counties and given each predictor X_{lk}, let the correlation within schools be $Corr(Y_{ijl}|X_{lk}, Y_{i'jl}|X_{lk}) = \rho_k^w, i \neq i'$, and let the correlation between schools within county be $Corr(Y_{ijl}|X_{lk}, Y_{i'j'l}|X_{lk}) = \rho_k^b, j \neq j'$. Data can be re-structured in a similar way as (2.10) where a vector of an outcome is replicated twice at the county level and the re-arranged covariate matrix has a block diagonal structure. To account for the correlations within schools and the correlation between schools in a county, a three-level exchangeable working covariance matrix can be expressed as

$$\tilde{\mathbf{V}}_l = \begin{bmatrix} \sigma_1^2\mathbf{R}_1 & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{R}_2 \end{bmatrix}, \mathbf{R}_k = \begin{bmatrix} ex(\rho_k^w) & \rho_k^b\mathbf{1}_{n_{jl}\times n_{jl}} & \cdots & \rho_k^b\mathbf{1}_{n_{jl}\times n_{jl}} \\ \rho_k^b\mathbf{1}_{n_{jl}\times n_{jl}} & ex(\rho_k^w) & \cdots & \rho_k^b\mathbf{1}_{n_{jl}\times n_{jl}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k^b\mathbf{1}_{n_{jl}\times n_{jl}} & \rho_k^b\mathbf{1}_{n_{jl}\times n_{jl}} & \cdots & ex(\rho_k^w) \end{bmatrix}_{n_{jl}n_l\times n_{jl}n_l}, k = 1,2.$$

where $ex(\rho_k^w)$ is a $n_{jl}\times n_{jl}$ exchangeable correlation structure with the correlation ρ_k^w. The matrix $\mathbf{1}_{n_{jl}\times n_{jl}}$ is a $n_{jl}\times n_{jl}$ one matrix, and the dispersion parameter or variance parameters are σ_k^2.

2.4 Simulation Study

We conducted a small scale simulation study to provide guidance on practical approaches to estimate model parameters and to examine properties of estimators and hypothesis tests. In available software, the most straightforward way to implement the proposed method is to use the working independence assumption within cluster.

In standard GEEs, this approach is fully efficient when cluster sizes are equal and covariates are invariant or mean-balanced within cluster, but can suffer severe efficiency loss otherwise (*Mancl and Leroux*, 1996).

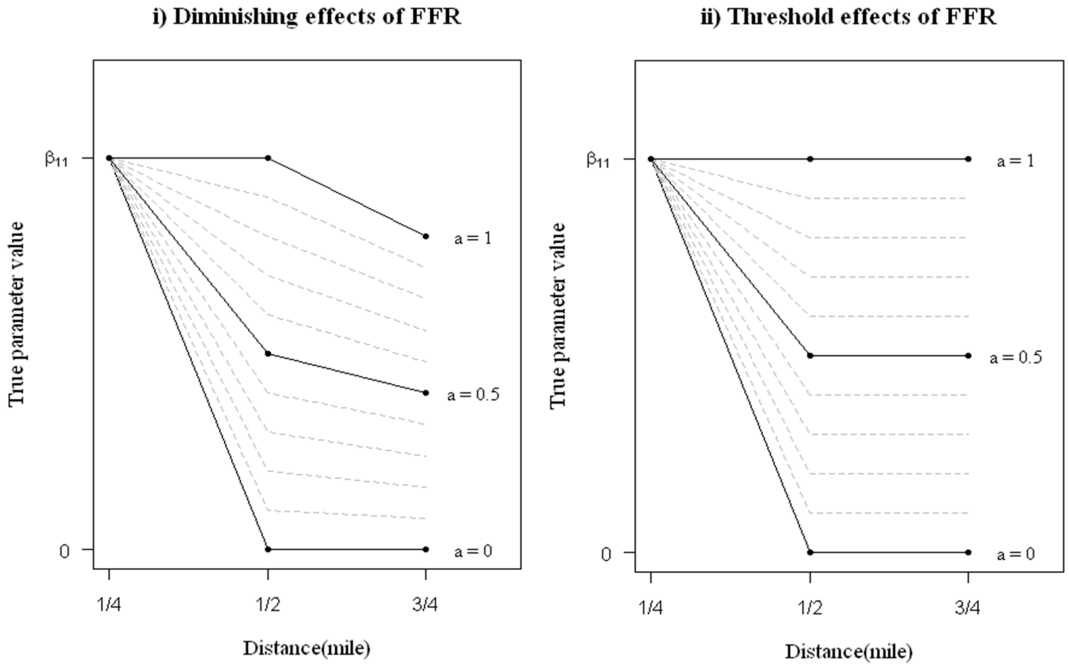
Because the environmental effects in our motivating study will typically be small, loss or gain in efficiency may have important implication for derived inferences. Hence, we examine statistical power of detecting a small degree of differences of environmental effects assuming unequal large cluster sizes and invariant covariates within cluster.

2.4.1 Simulation Setup

We set up the simulations to reflect two possible scenarios of the comparison of the marginal effects of fast food restaurants (FFR) across several buffers on child's BMIz: 1) marginal effects of FFR are diminishing with distance and 2) marginal effects of FFR have threshold at some distance. For both simulation scenarios, sample size, nesting structure (i.e., number of clusters and subjects per cluster), and distribution of the multiple informants (the number of restaurants within $1/4, 1/2$, and $3/4$ miles from each school) were the same as observed in the data example. For instance, in our motivating data the average number of children per school and its standard deviation are 145.6 and 159.5, respectively, yielding a coefficient of variation (CV) of 1.1. This means that unbalance of cluster size is large. For each simulation scenario we simulated 1,000 datasets where each data contain 926,018 observations nested in 6,323 clusters. Multiple informant predictors were fixed to the observed number of FFR in the motivating example (see Table 2.1), thus we only generated outcome data conditional on the predictors $\text{FFR}_k, k = 1, 2, 3$.

To simulate data with diminishing effects of FFR with distance, the marginal effect of FFR on child's BMIz within $1/4$ miles was fixed at the observed value in our motivating example (β_{11}=0.0234). True parameters of FFR within distance $1/2$ and $3/4$ miles, β_{12} and β_{13}, were set to $\beta_{12} = a\beta_{11}$ and $\beta_{13} = 0.8a\beta_{11}$, where $0 \leq a \leq 1$, i.e.,

Figure 2.1: **Regression parameter values used in simulations for i) diminishing and ii) threshold effects of fast food restaurants (FFR) at distance 1/4, 1/2, and 3/4 miles. True parameter settings are $\beta_{12} = a\beta_{11}$ at 1/2 mile, $\beta_{13} = 0.8a\beta_{11}$ at 3/4 mile for diminishing effects and $\beta_{12} = a\beta_{11}$ at 1/2 mile, $\beta_{13} = a\beta_{11}$ at 3/4 mile for threshold effects, such that a $(0 \leq a \leq 1)$ controls the differences across parameters.**



the effects of FFR consistently decrease over distance. Similarly, for threshold effects with distance, $\beta_{12} = a\beta_{11}$ and $\beta_{13} = a\beta_{11}$, where $0 \leq a \leq 1$, i.e., the effects of FFR decrease at some distance and continue to be constant. Here, the constant a controls the differences across regression parameters. Figure 2.1 shows regression parameter values used in the simulations for a range of values of a, for both diminishing effects and threshold effects.

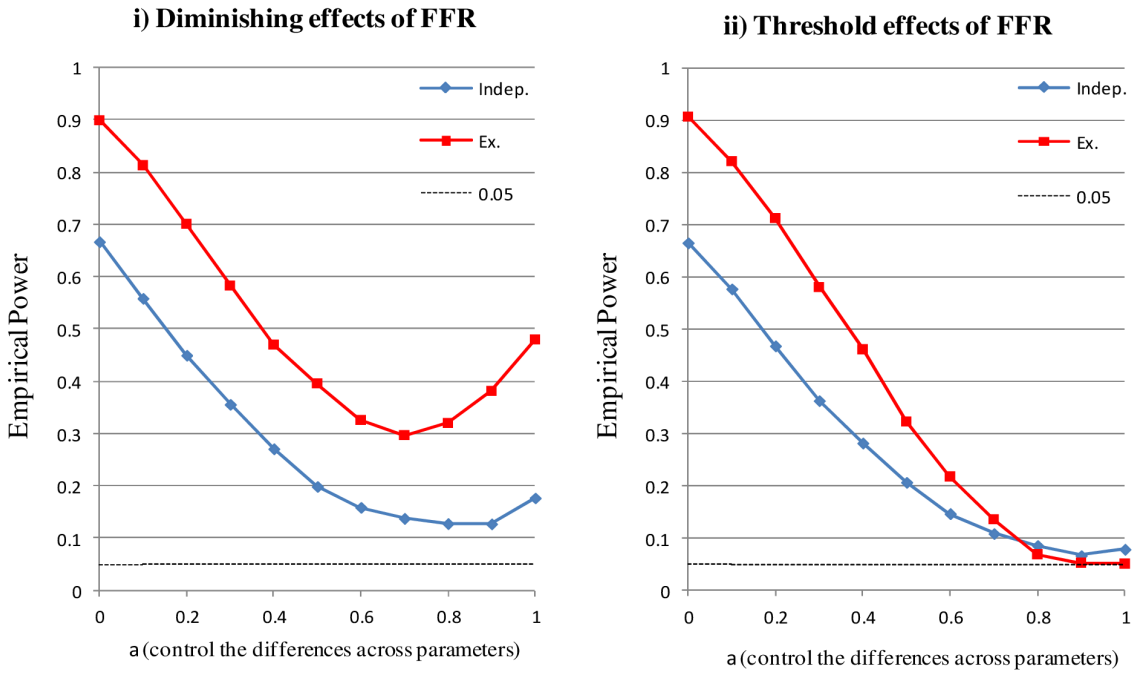
Note that, given the observed variances of the predictors (Table 2.1), these effects constrain the marginal covariances between predictors and outcome to $\sigma_{X_1Y} = \beta_{11}\sigma^2_{X_1} = 0.011, \sigma_{X_2Y} = a\beta_{11}\sigma^2_{X_2}$ and $\sigma_{X_3Y} = a\beta_{11}\sigma^2_{X_3}$ for threshold effects, and $\sigma_{X_3Y} = 0.8a\beta_{11}\sigma^2_{X_3}$ for diminishing effects. Further we assumed the marginal outcome mean was $\mu_Y = 0$ (centered), and had marginal variance $\sigma^2_Y = 0.14$ (as observed in our motivating data).

To simulate outcomes, we first generated cluster level values from a normal distribution with mean $E[\bar{Y}_{\cdot j}|\text{FFR}_{j1}, \text{FFR}_{j2}, \text{FFR}_{j3}] = \gamma_0 + \gamma_1\text{FFR}_{j1} + \gamma_2\text{FFR}_{j2} + \gamma_3\text{FFR}_{j3}$ and variance $Var[\bar{Y}_{\cdot j}|\text{FFR}_{j1}, \text{FFR}_{j2}, \text{FFR}_{j3}] = \sigma^2_\delta$, where $\sigma^2_\delta = \sigma^2_Y - \gamma^2_1 Var(\text{FFR}_{j1}) - \gamma^2_2 Var(\text{FFR}_{j2}) - \gamma^2_3 Var(\text{FFR}_{j3}) - 2\gamma_1\gamma_2 Cov(\text{FFR}_{j1}, \text{FFR}_{j2}) - 2\gamma_2\gamma_3 Cov(\text{FFR}_{j2}, \text{FFR}_{j3}) - 2\gamma_3\gamma_1 Cov(\text{FFR}_{j3}, \text{FFR}_{j1})$. The conditional mean of the cluster level given all three predictors was used because the outcome needs to be simulated only once given all three predictors. We used the Sweep operator (*Beaton*, 1964; *Dempster*, 1969) to derive conditional associations, $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ given the specified marginal associations $\beta_{11}, \beta_{12}, \beta_{13}$ (see Appendix B). Given the cluster mean, we generated subject level observations as $Y_{ij} = \bar{Y}_{\cdot j} + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$, where $\sigma^2_\epsilon = (1 - \rho_y)/\rho_y\sigma^2_\delta$, and $\rho_y = Corr(Y_{ij}, Y_{i'j})$ set to 0.05 for $i \neq i'$ in the j^{th} cluster. With this parameter setting, the true marginal covariance matrix for the HMIM (2.11) has non-equal blocks (see Appendix C).

2.4.2 Simulation Results

Let $\hat{\boldsymbol{\beta}}_{Ex}$ denote the estimator for $\boldsymbol{\beta}$ in (2.12) when using $\mathbf{V}_{jk} = \phi_k\mathbf{R}_{jk}$ as the diagonal blocks of $\mathbf{V}_j$ (2.11), with $\mathbf{R}_{jk}$ being an exchangeable correlation structure with parameter $\rho_k, k = 1, 2, 3$. Similarly, let $\hat{\boldsymbol{\beta}}_I$ denote the estimator for $\boldsymbol{\beta}$ when $\mathbf{V}_j$ consists of blocks of $\mathbf{V}_{jk} = \phi_k\mathbf{R}_{jk}$ with $\mathbf{R}_{jk}$ being an independence correlation structure. From the 1,000 datasets, the empirical power was calculated as the rate of rejecting the overall test for comparing marginal effects β_{1k} for $k = 1, 2, 3$, i.e., $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ vs. H_1 : at least one differs, for both estimators. For a given data set, the null hypothesis was rejected when the Wald test statistic T (see Section 2.3.2) exceeded the critical value for a chi-squared distribution with 2 degrees of freedom. As shown in Figure 2.2, the empirical power of $\hat{\boldsymbol{\beta}}_{Ex}$ was uniformly higher than $\hat{\boldsymbol{\beta}}_I$ for diminishing effects of FFR. Power was always greater than the significance level (0.05) because in the range of a ($0 \leq a \leq 1$) true parameters of FFR were always

Figure 2.2: **Simulation results assessing power for the hypothesis test $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ for i) diminishing and ii) threshold effects of fast food restaurants (FFR) using HMIM with exchangeable (Ex.) and independence (Indep.) correlation structures. True parameter settings are $\beta_{12} = a\beta_{11}, \beta_{13} = 0.8a\beta_{11}$ for diminishing effects and $\beta_{12} = a\beta_{11}, \beta_{13} = a\beta_{11}$ for threshold effects.**



distinguishable. The U-shape of the power function within the range of a is due to the non-centrality parameter of the test statistic, T, being a quadratic function of a under the alternative hypothesis.

For the threshold effects of FFR, the empirical powers of both estimators $\hat{\boldsymbol{\beta}}_{Ex}$ and $\hat{\boldsymbol{\beta}}_{I}$ go to nominal value (0.05) of Type I error rate when $a = 1$, or $\beta_{11} = \beta_{12} = \beta_{13}$. When β_{1k}s are distinguishable or a goes to 0, the power for $\hat{\boldsymbol{\beta}}_{Ex}$ increases faster than for $\hat{\boldsymbol{\beta}}_{I}$. The crossing of the power curves of the estimators may be due to Monte Carlo errors from the simulation. The Monte Carlo errors will be negligible with an increased number of simulations.

Because of the large number of clusters, the empirical power from the overall test when $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ is true preserved 5% Type I error rate. When the number

of clusters is small, a bias corrected sandwich estimator could be used (*Mancl and Derouen*, 2001).

2.4.3 Simulation conclusions

Accounting for correlation within cluster is important to better detect small differences between marginal effects in an environmental study of clustered or hierarchical data. For instance, for threshold and diminishing effects, this simulation shows that, if using $\hat{\boldsymbol{\beta}}_{Ex}$, 80% power was achieved when $a < 0.2$. That is, to be statistically distinguishable, the association between the outcome and number of FFR at the outer buffers needs to be at most 20% of the association with the number of FFR in the inner buffer. However, note that if using $\hat{\boldsymbol{\beta}}_I$, 80% power could not be reached for any value of a.

The fact that power using $\hat{\boldsymbol{\beta}}_{Ex}$ is higher than $\hat{\boldsymbol{\beta}}_I$ can be explained by applying previous work on asymptotic relative efficiency (ARE) of *Mancl and Leroux* (1996). According to their formula for ARE, and given that we have the conditions: 1) invariant covariates within clusters, 2) unequal cluster sizes (CV ≈ 1.1), 3) large cluster size ($\bar{J} = 145.6$), and 4) intra-cluster correlation $\rho_k \approx 0.05, k = 1, 2, 3$, the ARE of $\hat{\boldsymbol{\beta}}_{Ex}$ to $\hat{\boldsymbol{\beta}}_I$ in the current data is about 0.55, meaning approximately 45% loss of efficiency by employing the independence correlation structure even for small intra-cluster correlation and invariant covariates within clusters.

This simulation study shows that a HMIM should be employed for formal testing of associations of an outcome among correlated predictors in clustered or hierarchical data to increase power.

2.5 Data Example

We used data for children who participated in the 2007 California physical fitness test (also known as FitnessGram), which contains direct measures of children's

weight and height, among all children attending $5^{th}, 7^{th}$ and 9^{th} grade, as well as other covariates such as age, sex and race. Following prior exclusion criteria, we used data on 926,018 children nested in 6,323 schools (*Sánchez et al.*, 2012). The location of fast food restaurants (FFR) and convenience stores (CS) in California was purchased from InfoUSA, a commercial source. Geocodes for schools and food stores were cross-referenced to obtain the counts of stores within $1/4, 1/2, 3/4$ miles from schools, denoted by $\text{FFR}_1, \text{FFR}_2, \text{FFR}_3$ and $\text{CS}_1, \text{CS}_2, \text{CS}_3$. We obtained data from the California Department of Education's databases and the 2000 US Census to characterize the size and racial/ethnic composition of the schools, as well as the socio-economic conditions of the neighborhoods in which schools were located.

Body mass index z-score (BMIz) was used as a continuous outcome. BMIz was derived by calculating body mass index (weight in kg/height in meters squared), and standardizing it according to an age and gender-specific BMI distribution. In other words, BMIz indicates how much a child's BMI differs from a reference group of the same age and gender (*CDC*, 2005). In contrast to BMI among adults, BMI among children needs to be standardized to a reference population because they are still growing and their body composition is changing as they grow (*Must and Anderson*, 2006) such that the meaning of BMI is not the same across age and sex. Following prior analyses (*Sánchez et al.*, 2012), we included individual- and school-level covariates as adjustment factors in models. The individual-level covariates are grade, age, gender, race/ethnicity. The school-level covariates are school's racial composition, school's neighborhood-level education, school's total enrollment, and percent of children enrolled in the free or reduced price meal program.

Descriptive statistics of child's BMIz, the number of FFR and CS within distance $1/4, 1/2, 3/4$ miles are summarized in Table 2.1. The average number of children per school and its standard deviation are 145.6 and 159.5, respectively, yielding a coefficient of variation (CV) of 1.1.

Table 2.1: **Descriptive statistics for BMIz*, number of fast food restaurants (FFR) and convenience stores (CS) at three distances, and their pairwise correlations.**

Distance	Variable	Mean	SD	Corr.	BMIz*	FFR_1	FFR_2	FFR_3	CS_1	CS_2	CS_3
	BMIz*	0.744	0.374	BMIz*	1	0.04	0.09	0.13	0.13	0.20	0.25
1/4 mile	FFR_1	0.233	0.679	FFR_1		1	0.55	0.36	0.25	0.23	0.21
1/2 mile	FFR_2	1.150	1.758	FFR_2			1	0.73	0.19	0.38	0.37
3/4 mile	FFR_3	2.709	2.915	FFR_3				1	0.15	0.37	0.50
1/4 mile	CS_1	0.178	0.473	CS_1					1	0.53	0.39
1/2 mile	CS_2	0.774	1.083	CS_2						1	0.76
3/4 mile	CS_3	1.697	1.845	CS_3							1

BMIz* is the mean of child's BMIz within schools

We conducted two sets of analyses: 1) the comparison of two different features of food stores within the same buffer, and 2) the comparison of a food environment feature across several buffer sizes. In both sets of analyses we fitted a HMIM with both exchangeable and independence structures and, for comparison, also MIM without accounting for cluster correlation. Further, the individual- and school-level covariates described above were included.

First, for the comparison of two different features of food stores within the same buffer, the counts of FFR and CS were standardized to a mean of zero and a standard deviation of one because of potentially different scales (e.g., an overall preponderance of one feature may be different compared to the other) so that the coefficients are in standard deviation units. We use $F^s_{j1}, F^s_{j2}, F^s_{j3}$ and $C^s_{j1}, C^s_{j2}, C^s_{j3}$ to denote the standardized number of FFR and CS within $1/4, 1/2, 3/4$ miles from the j^{th} school, respectively, and $\mathbf{Z}_{ij}$ to denote the vector of the individual- and the school-level covariates or confounders. For each of three buffers, $k = 1, 2, 3$, the fitted models are

$$E[\text{BMIz}_{ij}|F^s_{jk}, \mathbf{Z}_{ij}] = \beta^{F^s}_{0k} + \beta^{F^s}_{1k} F^s_{jk} + \mathbf{Z}^T_{ij}\boldsymbol{\beta}^{F^s}_{(k)},$$
$$E[\text{BMIz}_{ij}|C^s_{jk}, \mathbf{Z}_{ij}] = \beta^{C^s}_{0k} + \beta^{C^s}_{1k} C^s_{jk} + \mathbf{Z}^T_{ij}\boldsymbol{\beta}^{C^s}_{(k)}.$$

The null hypotheses of interest are whether for each buffer, the association between the number of CS and BMIz is the same as the association between the number of

Table 2.2: **Estimated associations* of two different features of the food environment (fast food restaurants vs. convenience stores) within the same buffer on BMIz and hypotheses tests of equality of the associations. Associations are estimated based on standardized number of fast food restaurants and convenience stores, adjusting for individual- and school-level covariates, using the proposed HMIM and the MIM without accounting for within cluster correlation.**

		HMIM				MIM	
		Exchangeable		Independence			
Distance	Association* or Test	Est.	(SE)	Est.	(SE)	Est.	(SE)
1/4 mile	Std. Fast Food (β_{11}^{Fs})	0.77	(2.54)	-0.11	(2.33)	-0.11	(0.99)
	Std. Convenience Stores (β_{11}^{Cs})	10.82	(2.44)	7.78	(2.40)	7.78	(1.12)
	H_0: $\beta_{11}^{Fs} = \beta_{11}^{Cs}$	P = 0.001		P = 0.009		P < 0.0001	
1/2 mile	Std. Fast Food (β_{12}^{Fs})	1.96	(3.46)	2.75	(2.66)	2.75	(1.10)
	Std. Convenience Stores (β_{12}^{Cs})	11.56	(2.88)	10.35	(2.72)	10.35	(1.14)
	H_0: $\beta_{12}^{Fs} = \beta_{12}^{Cs}$	P = 0.003		P = 0.009		P < 0.0001	
3/4 mile	Std. Fast Food (β_{13}^{Fs})	5.02	(3.36)	5.59	(2.73)	5.59	(1.15)
	Std. Convenience Stores (β_{13}^{Cs})	13.24	(3.00)	13.80	(2.82)	13.80	(1.18)
	H_0: $\beta_{13}^{Fs} = \beta_{13}^{Cs}$	P = 0.005		P = 0.004		P < 0.0001	

* Estimate and SE were multiplied by 10^3 to enhance readability

FFR and BMIz, i.e., $H_0 : \beta_{1k}^{Fs} = \beta_{1k}^{Cs}$ for $k = 1, 2, 3$.

Table 2.2 provides the results for the comparison of FFR and CS within the same buffer size. For all buffer sizes, the adjusted associations of CS with BMIz are significantly greater than those of FFR with BMIz. For example, given the 1/4 mile buffer size, child's BMIz increases 0.77×10^{-3} and 10.82×10^{-3} per one standard deviation increase of FFR (= 0.679) and CS (=0.473), respectively, after adjusting for individual and school factors. Using either an exchangeable correlation structure or the independence structure, we reached the same substantive conclusion, although the point estimates are slightly different. Second, to investigate how the association between the number of FFR (or CS) and BMIz varies across several buffers, we fitted

models

$$E[\text{BMIz}_{ij}|\text{FFR}_{jk}, \mathbf{Z}_{ij}] = \beta_{0k} + \beta_{1k}\text{FFR}_{jk} + \mathbf{Z}_{ij}^T\boldsymbol{\beta}_{(k)} \text{ for } k = 1, 2, 3 \text{ (similar for CS)}.$$

The question of interest is whether the associations between the number of a given foods store type (FFR or CS) varies across buffer sizes, i.e., the overall null hypothesis $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$. Note that the coefficients are expressed in units of BMIz per one unit increase in the number of stores, since the same feature is being compared across buffers. The parameter estimates and the p-values for the overall hypotheses tests are given in Table 2.3.

The overall test for the associations of FFR across several buffers is not rejected (p=0.895), meaning the number of FFR within $1/4, 1/2$, and $3/4$ mile from schools do not have significantly different associations with child's BMIz. This implies that the most relevant buffer size is $3/4$ mile from schools, or potentially further. Child's BMIz increases 1.72×10^{-3} per one FFR increment within $3/4$ miles from schools (p=0.862, not reported in Table 2.3). By employing either an exchangeable correlation matrix or the independence structure, the same conclusion is derived.

Unlike the associations of FFR, the associations of CS across the buffers are significantly different based on HMIM with the exchangeable correlation matrix ($p = 0.004$). Based on the result of HMIM with the exchangeable correlation matrix, we performed subsequent hypothesis test as described in Section 2.3.2. i.e., test the one-sided null hypothesis, $H_0 : \beta_{11} \leq \beta_{12}$. We rejected the one-side null hypothesis (p-value = 0.017) and concluded that the most relevant buffer size for the association between CS and child's BMIz is $1/4$ mile from schools. Child's BMIz increases 0.022 per one CS increment within $1/4$ miles from schools after adjusting other covariates. Note that in Table 2.2 and Table 2.3 the estimates and standard errors using an exchangeable correlation matrix differ from those using the independence assumption. These

Table 2.3: **Estimated associations* between number of fast food restaurants (or convenience stores) and BMIz across three distances and test of equality of association across distances. Associations are estimated from three different models, adjusting for individual- and school-level covariates.**

		HMIM				MIM	
		Exchangeable		Independence			
Distance	Association* or Test	Est	(SE)	Est	(SE)	Est	(SE)
1/4 mile	Fast Food (β_{11})	1.14	(3.73)	-0.17	(3.21)	-0.17	(1.36)
1/2 mile	Fast Food (β_{12})	1.12	(1.97)	1.56	(1.42)	1.56	(0.59)
3/4 mile	Fast Food (β_{13})	1.72	(1.15)	1.92	(0.88)	1.92	(0.37)
$H_0: \beta_{11} = \beta_{12} = \beta_{13}$		P = 0.895		P = 0.782		P = 0.265	
1/4 mile	Convenience store (β_{11})	22.89	(5.16)	16.47	(4.77)	16.47	(2.23)
1/2 mile	Convenience store (β_{12})	10.67	(2.66)	9.55	(2.36)	9.55	(0.99)
3/4 mile	Convenience store (β_{13})	7.17	(1.62)	7.48	(1.44)	7.48	(0.60)
$H_0: \beta_{11} = \beta_{12} = \beta_{13}$		P = 0.004		P = 0.118		P = <0.001	

* Estimate and SE was multiplied by 10^3 to enhance readability

changes result in test statistics (e.g., a ratio of the difference between two regression parameters to its standard error) that are larger when using the exchangeable correlation matrix and thus smaller p-values. For instance, in Table 2.3 the overall null hypothesis yields a p-value = 0.118 when using the independence assumption within cluster while an exchangeable assumption yielded a p-value = 0.004, highlighting the gain in efficiency when using the exchangeable vs. independence assumption.

Lastly, as shown in Table 2.2 and Table 2.3, the MIM without accounting for within-cluster correlation provides the same point estimates of HMIM with the independence structure, but the failure to account for hierarchical structures yields the underestimated standard errors, resulting in invalid inference.

2.6 Discussion

We extended multiple informant methods to a hierarchical data setting to enable comparison of the associations between multiple correlated predictors on a univariate outcome measured in clustered sets of individuals. The method is based on a

non-standard application of generalized estimating equations and can be applied to settings where the outcome is continuous, count, or binary. In simulation study, we showed the improved power and efficiency of estimators based on using a block diagonal of exchangeable correlation matrices instead of the working independence correlation structure. A practical advantage of a HMIM is that it can be fitted using available GEE software. A marginal GEE model for each multiple informant can be separately fitted, and then a joint empirical variance estimator can be calculated to conduct hypothesis tests involving the marginal effects of predictors. We applied HMIMs to examine how the association between the number of fast food restaurants (or convenience stores) and child's BMIz varies across several buffers from schools and to compare the association of two different features of the food environment (fast food restaurants vs. convenience stores) with child's BMIz. The overall hypothesis that the association between number of FFR and child's BMIz across several buffers is the same was not rejected, suggesting that the association of the count of FFR up to 3/4 mile from schools does not differ significantly from the association at small buffer sizes with accounting for individual- and school-level covariates. In contrast, the association of BMIz with the count of CS differs depending on distance from schools, with 1/4 mile being most relevant buffer size. We also showed the association between the count of CS and BMIz is much stronger compared to the association between FFR and BMIz.

We proposed a testing strategy that may be helpful in selecting an appropriate buffer size at which to estimate the association between an environmental feature and an outcome. However, there are some extreme cases where the hypothesis testing strategy may fail. For instance, if there are few or no additional of food stores between distances r_{k-1} and r_k, then the marginal association between the number of food stores and child's BMIz across buffers r_{k-1} and r_k would likely to be the same due to $X_{j,k-1} \approx X_{jk}$. An alternative might be to define multiple informant covariates as

the new information not contained in the previous buffer, e.g., $X_j(\delta_k) = X_{jk} - X_{j,k-1}$, and fit a model $g(\mu_{ij}) = \beta_0^* + \beta_1^* X_j(\delta_1) + \beta_2^* X_j(\delta_2) + \beta_3^* X_j(\delta_3)$. If $\beta_3^* \neq 0$, then the buffer of size is r_3 from schools still provides information on child's BMIz, and similarly for the other buffer sizes. However, the interpretation of these coefficients is that of conditional associations, not marginal associations.

We also confirmed the underestimated variances of the estimators from the MIM due to the failure to incorporate hierarchical structures, which provide us invalid inference. The bootstrap method, as pointed out by a referee, may be employed for valid inference. For instance, suppose that we have 5,000 bootstrap estimates of regression parameters from the MIM. Then, the empirical variance/covariance of the estimates can be used for hypothesis testing because the failure to account for hierarchical structures has little impact on the population point parameter estimates. The bootstrap method, however, may require extensive computational time.

The main idea of MIM is very similar to seemingly unrelated regression methods (SUR) (*Zellner*, 1962). The main difference between MIM and SUR is that in MIM the same outcome is replicated to form an outcome vector for the cluster with predictors changing from one replicate to another, whereas different outcomes form an outcome vector in SUR. Similarly, a HMIM is related to the model structure described by *Rochon* (1996). The author employed SUR in a repeated measures setting for discrete and continuous outcome variables; nevertheless, here we extended MIM to hierarchical data to enable estimation and testing of marginal effects of several correlated factors or multiple informants or predictors.

CHAPTER III

Distributed Lag Models: Examining Associations between the Built Environment and Health

3.1 Introduction

Studying the contributions of both individual- and built environment factors on health is important to better understand the determinants of disease because environmental factors may directly constrain individual's behaviors and choices (*Diez-Roux*, 1998; *Susser*, 1994). Built environment factors near or around schools, in particular commercial establishments offering "junk" food have recently received particular attention as possible contributors to the childhood obesity epidemic. The availability of establishments that sell high energy, low nutrition foods near schools may increase consumption of these items, both through purchasing and consumption on the way to and from school, and indirectly through excess exposure to advertising that may shape children's dietary choices and weight status (*Gebauer and Laska*, 2011; *Hillier et al.*, 2009). *Davis and Carpenter* (2009) showed that children's obesity status was associated with proximity of fast food restaurants to schools. Similarly, *Sánchez et al.* (2012) showed that the number of convenience stores within a 1/2 mile radius from a school (also known as a 1/2 mile "buffer") was significantly associated with children's obesity status although not all studies show consistent results (*Alviola et al.*, 2014;

Currie et al., 2009; *Harris et al.*, 2011; *Langellier*, 2012). Inconsistent findings may be in part due to how measures of the built environment are constructed; for instance, the buffer size within which features of the environment are counted.

Studies relating food store availability near schools to children's weight typically choose buffer sizes on an ad hoc manner (e.g., 1/4, 1/2, or 1 miles from schools), often justifying these distances on the basis of the time it takes to walk such distances, e.g., children may walk a given distance in 5-10 minutes (*An and Sturm*, 2012; *Howard et al.*, 2011; *Zenk and Powell*, 2008). To determine the most influential buffer size given a-priori selected buffer sizes, estimated associations are compared by examining the extent of overlap of the corresponding confidence intervals; or by examining the distance at which the associations are, or cease to be, significant (*Davis and Carpenter*, 2009).

The distances within which the presence of certain food stores significantly affect children's weight remain unknown and statistical methods to determine them empirically have been understudied. Although some approaches have been proposed to empirically select the most relevant/influential/appropriate distances to health outcomes, there is currently no consensus on the most robust method. In the transport geography literature, *Guo and Bhat* (2004) proposed using goodness-of-fit statistics comparing several fitted models as a way to empirically select the best buffer size. Conversely, *Spielman and Yoo* (2009) demonstrated through a simulation study that selecting buffer sizes based on goodness-of-fit statistics does not perform well and may even result in biased associations with the health outcome.

Indeed, misspecifying the buffer size within which built environment features are measured may yield biased associations and incorrect inference. This issue is more generally known as the modifiable area unit problem (MAUP) (*Fotheringham and Wong*, 1991; *Openshaw*, 1996). For instance, consider data generated from the model $Y = \beta X(A_5) + \epsilon$, where $\beta \neq 0, X(A_5)$ indicates a spatially correlated built environ-

ment factor measured within a buffer of radius 5 from locations of interest, while ϵ is a residual error. Without knowledge about the true buffer size, suppose we instead fit $Y = \theta_0 + \theta_1 X(A_3) + \epsilon'$. Since the built environment feature located between distance 3 to 5 from the locations, say $X(A_{3-5})$ is correlated with $X(A_3)$, the estimated θ_1 is biased; i.e., $X(A_{3-5})$ confounds the association between Y and $X(A_3)$ rendering the estimate for θ_1 different from $\beta(\theta_1 \neq \beta)$.

To better understand the associations between built environment features and health as a function of distance from the locations of interest, we propose using distributed lag models (DLMs) and apply DLMs to examine the effects of the food environment around schools on children's body mass index z-score (BMIz).

DLMs have a long history in economics research mostly derived from the works of *Koyck* (1954), *Nerlove* (1956), *Almon* (1965) and others; more recently they have been used in air pollution studies (*Braga et al.*, 2014; *Dominici et al.*, 2004; *Goodman et al.*, 2004; *Heaton and Peng*, 2012; *Pope and Schwartz*, 1996; *Pope et al.*, 1991; *Welty et al.*, 2009; *Zanobetti et al.*, 2000) to examine associations between lagged exposure covariates on an outcome. In application to air pollution studies, the lagged exposure used in DLMs is the air pollution in the previous L days. In our built environment research, the lagged exposure covariate is represented by the number of food stores between two radii, r_{l-1} and $r_l, l = 1, 2, \ldots, L$, where $r_0 = 0$, away from schools. In other words, we consider the number of food stores within "ring"-shaped areas around schools and examine the association between these lagged exposure covariates on children's weight.

Since the associations between a health outcome and the built environment features within ring-shaped areas are likely to be similar in adjacent rings, we model them as a smooth function of distance from the locations of interest, in this case schools. While various ways for constraining the lagged coefficients of DLMs have been proposed in the literature (*Heaton and Peng*, 2012; *Welty et al.*, 2009; *Pope and*

Schwartz, 1996), here we follow *Zanobetti et al.* (2000) and use splines to model the DL coefficients. Smoothing the DL coefficients has the advantage of stabilizing the estimates of the coefficients and their variances, and enables researchers to use a large number L of lags without necessarily increasing the degrees of freedom of the model.

In the following sections of the paper, the text is organized as follows: In Section 3.2, we introduce DLMs in the context of studying the association between the built environment and health outcomes, including how associations of the built environment may vary according to subject characteristics. In Section 3.3, we (a) perform a small scale simulation study that aims to show how DLMs can capture true underlying associations between health and built environment features given various conditions of the built environment, and (b) compare them to currently used traditional regression methods. In Section 3.4, we apply the proposed method (DLMs) to examine the effect of availability of convenience stores (CS) on children's BMI z-score (BMIz) using a surveillance dataset for 5^{th} and 7^{th} grade children who attended public schools in the State of California. Section 3.5 concludes with a discussion.

3.2 Distributed Lag Model (DLM)

3.2.1 Statistical Model

Let Y_i be a continuous outcome measured at location i and $X_i(r_{l-1}; r_l), l = 1, 2, \ldots, L$, be a feature of the built environment measured within a ring-shaped area around location i with inner and outer radius r_{l-1} and r_l, respectively. In our example, schools are the unit of observations, the health outcome is the average BMIz of children attended in the school, and the built environment measure is the number of CS near the schools. We denote with r_L the maximum distance around locations beyond which we assume there is no further association between the environment feature and the outcome. The total number of lags L considered can be chosen large

enough to allow for more flexibility in the shapes of the associations of the aggregated environment feature and the outcome. Then, in a DLM, the outcome is modeled as

$$Y_i = \beta_0 + \Sigma_{l=1}^{L} \beta(r_{l-1}; r_l) X_i(r_{l-1}; r_l) + \epsilon_i, \tag{3.1}$$

where $\epsilon_i \sim N(0, \tau^2), \beta_0$ represents the intercept term, and $\beta(r_{l-1}; r_l)$ denotes the association of the environment feature measured between radius r_{l-1} and r_l around the locations and the outcome. In our motivating example, the coefficient $\beta(r_{l-1}; r_l)$ represents the difference in mean children's BMIz per one higher CS count in the area between radii r_{l-1} and r_l (see Figure 3.1a).

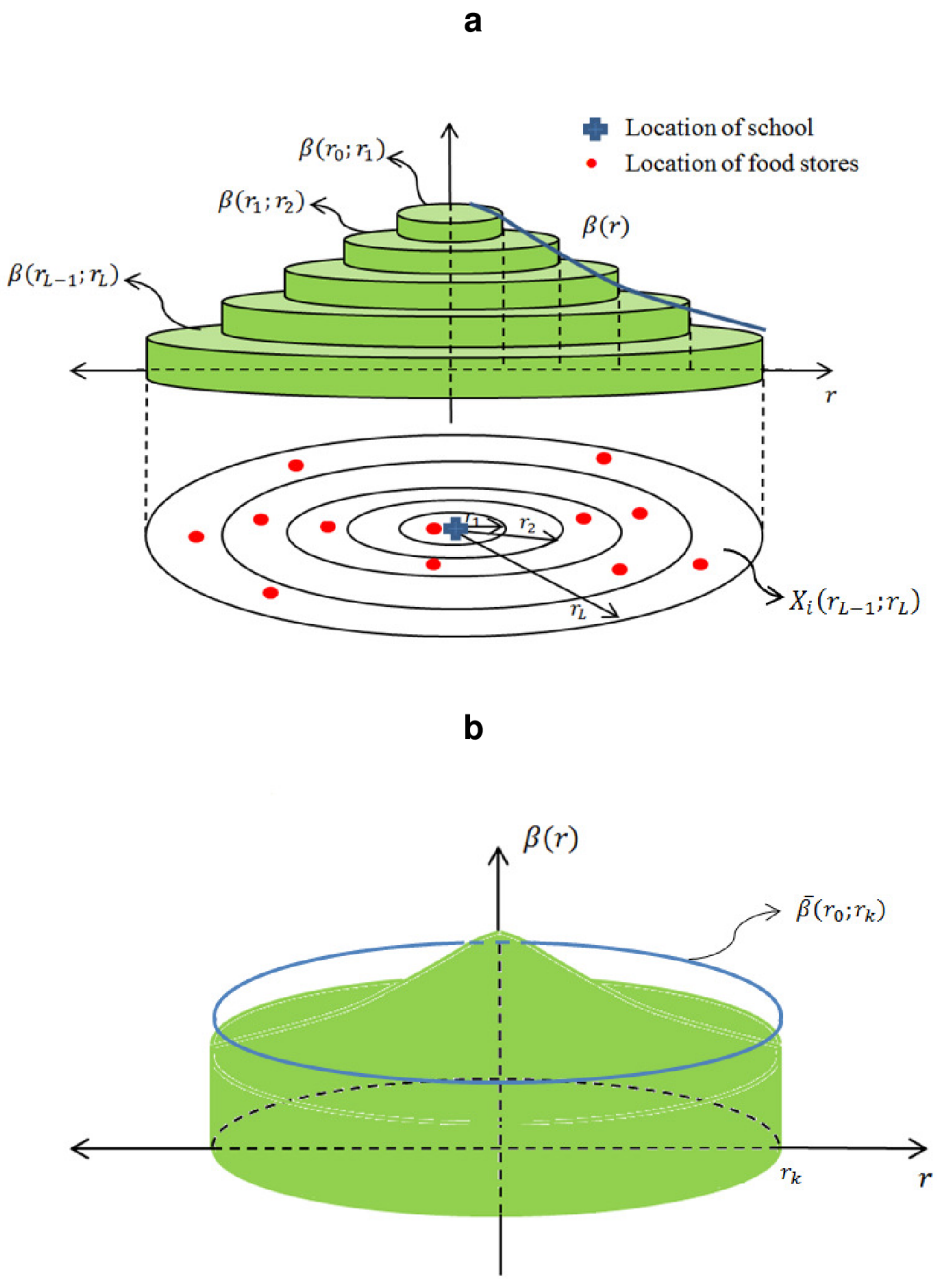
We constrain the coefficients $\beta(r_{l-1}; r_l)$ to vary as a smooth function of distance $r_l, l = 1, 2, \ldots, L$, by using splines (*Hastie and Tibshirani*, 1990; *Zanobetti et al.*, 2000). This ensures that coefficients corresponding to adjacent areas are similar, as we would not typically expect associations to change abruptly across distance. It also alleviates possible numerical problems that may arise when many locations have zero food stores between two given radii r_{l-1} and r_l . In particular, we model the association coefficients $\beta(r_{l-1}; r_l)$ using a radial basis function; that is

$$\beta(r_{l-1}; r_l) = \alpha_0 + \alpha_1 r_l + \Sigma_{k=1}^{L} \tilde{\alpha}_k |r_l - r_k|^3, \tag{3.2}$$

where α_0 denotes the intercept of the lag effects, α_1 represents the average change rate of lag association, and the $\tilde{\alpha}_k$ are penalized coefficients to achieve smoothness (see Appendix for details). While other types of splines could be employed, using smoothing splines, avoids the issue of knot selection.

Generally only interpreting the coefficient $\beta(r_{l-1}; r_l)$, rather than α's, will be of interest because it directly informs how the built environment feature captured by $X_i(r_{l-1}; r_l)$ is associated with the outcome. The units of the built environment feature captured by $X_i(r_{l-1}; r_l)$ naturally impact the interpretation of $\beta(r_{l-1}; r_l)$. In our

Figure 3.1: **(a) Ring-shaped areas within which food environment features are ascertained and corresponding DL coefficients. (b) Averaged coefficient associated with features within buffer of radius r_k, $\bar{\beta}(0; r_k)$.**



application, we used the total number of CS in ring-shaped areas, however we note that other definitions of $X_i(r_{l-1}; r_l)$ could be used, such as the density of CS per area. In that case, the DL coefficients could be readily calculated by transforming the parameters in Eq. (3.1): coefficients of the association between the density measure, i.e., the count per unit area $X_i(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)$, and health outcome are equal to the coefficients in Eq. (3.1) weighed by the area of the ring where $X_i(r_{l-1}; r_l)$ was calculated, i.e., $\beta(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)$.

Estimation of DLM parameters can be carried out using either a frequentist or a Bayesian approach: software to fit DLMs is available both in R (*R Development Core Team*, 2014; *Gasparrini*, 2011) and WinBugs (*Lunn et al.*, 2000). In fitting DLMs, one could estimate the smoothing parameter for the lag effects, i.e., the penalty parameter for $\tilde{\alpha}_k$, either by treating them as random effects or via generalized cross validation (GCV). We opted for random effects and fitted the model using a Bayesian approach (see appendix for details): this allows us to account for the uncertainty in the penalty parameters and easily derive the variance of model summaries such as the average buffer effects up to a certain distance $r_k, \bar{\beta}(0; r_k)$, defined below.

3.2.2 Connection between DLMs and traditional approaches

In environmental studies, traditional linear models of the form $Y_i = \beta X(0; r_k) + \epsilon_i$ is the widespread approach to estimate the average association between measures of the built environment within a buffer of radius r_k and a health outcome. An implicit assumption of such models is that the association between the outcome and the built environment factors within distance r_k is constant and no association beyond distance r_k exists. Our proposed DLM allows us to relax both of these assumptions. In addition, our model enables us to easily calculate the average buffer effect $\bar{\beta}(0; r_k)$ up to a given distance r_k (e.g., the average difference in children's BMIz per one additional food store within a buffer area of radius r_k) by comput-

ing the average height of the volume of the shape depicted in Figure 3.1b, i.e., $\bar{\beta}(0; r_k) = \Sigma_{l=1}^{k} \beta(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)/\pi r_k^2$. To see this, first consider the average buffer effect up to distance r_k in the density scale. As mentioned previously, the buffer effect between distance r_{l-1} and r_l in the density scale is the area-weighted association, $\beta(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)$ which is a volume of the area between two radii r_{l-1} and r_l. By summing the area-weighted associations, $\Sigma_{l=1}^{k} \beta(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)$, we can derive the average association within the buffer of radius r_k. Alternatively, if the density measure is used as covariate in the DLM, then the average association within the buffer of radius r_k is simply the sum of all coefficients up to distance r_k. We note that while in the air pollution literature, the simple sum of the DL coefficients represents the overall health impact of a unit difference in exposure on the previous k days, in this research of the built environment associations with child body weight, the DL coefficients have to be weighted by the area of the rings measured to obtain the equivalent interpretation.

We also note that deriving the average buffer effect up to distance r_k using the model in Eq. (3.1) provides a more accurate estimate of buffer effects than one could derive from the linear models traditionally used in environmental exposure studies. Our simulation studies presented in Section 3.3 will provide an illustration of this phenomenon.

3.2.3 Differences in DL coefficients by subject characteristics

As with any linear models, also DLMs could be expanded to allow the association between a health outcome and built environment features to vary by subject characteristics. For instance, effects between features of the built environment and children's BMIz might be different by age or grade because younger children may have more restrictions on what they are allowed to do within and outside of school boundaries (e.g., school policies on whether children are allowed to leave school for lunch may vary by

age). To investigate whether the DL effects vary according to subjects characteristics, Eq. (3.1) could include interaction terms between $X_i(r_{l-1}; r_l), l = 1, 2, \ldots, L$, and Z_i. Similarly to what is done previously, we would model the interaction coefficients using splines,

$$\theta(r_{l-1}; r_l) = v_0 + v_1 r_l + \Sigma_{k=1}^{L} \tilde{v}_k |r_l - r_k|^3, \tag{3.3}$$

where v_0 represents the difference in the intercept of the lag effects due to Z_i as compared to the baseline ($Z_i = 0$), and v_1 denotes the difference in the change rate over distance due to Z_i while the random coefficients $\tilde{v}_k$ are penalized coefficients used to achieve smoothness for the interaction term. Again, interpreting the effect of the interaction $\theta(r_{l-1}; r_l)$, rather than the v's, is generally of primary interest.

3.2.4 Extensions of the model

DLMs can be extended in several directions to allow for the examination of different types of outcomes. For example, we can define generalized DLMs if the observed outcome Y_i is binary or a count. In our motivating example, although our outcome is approximately normal, the assumption of constant variance, typical of linear models, does not hold. For this type of situations, instead of using a DLM, we can use a weighted DLM where the error terms ϵ_i are modeled as $\epsilon_i \sim N(0, \tau^2/w_i)$, where w_i is a known weight for the i^{th} observation. Fitting a weighted DLM is rather straightforward: we simply need to transform the data Y_i according to the following equation: $Y_i^w = Y_i\sqrt{w_i}$ and $X_i^w(r_{l-1}; r_l) = X_i(r_{l-1}; r_l)\sqrt{w_i}, l = 1, 2, \ldots, L$, and fit the usual DLM as given by (3.1) to the transformed data. Interpretation of the regression coefficients remains unchanged.

3.3 Simulation

We performed a small scale simulation study to improve our understanding of estimation and inference of associations of interest in DLMs are affected by the degree of spatial correlation in the built environment and the shape of association between measured environment factors and an outcome across distance. Further, we compared results obtained from DLMs to those obtained from traditional approaches based on linear models when the goal is to estimate the average association between features of the built environment and an outcome up to a-priori specified distances.

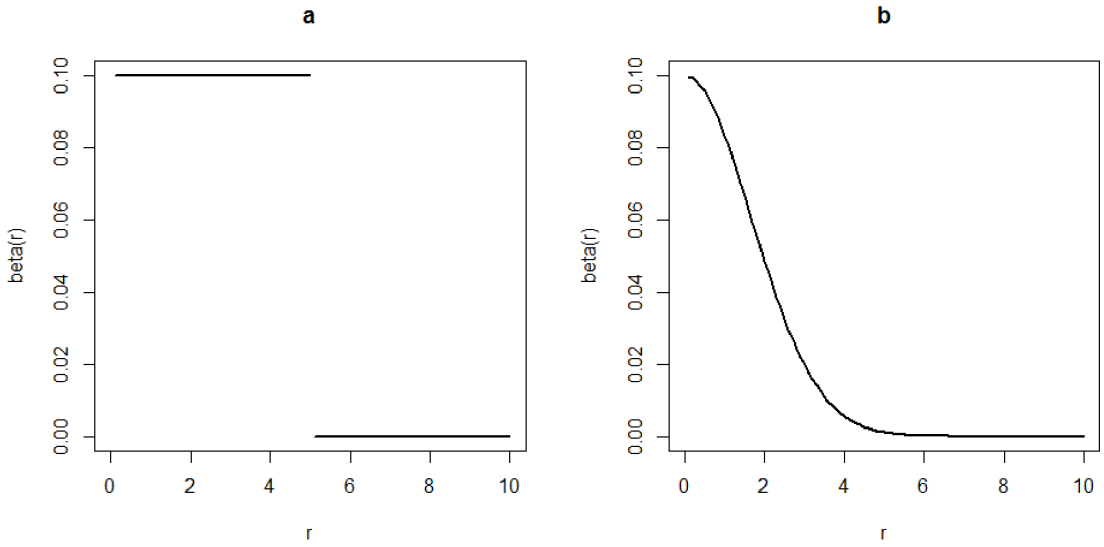
For our simulations, we used a spatial domain as the square $(0, 500) \times (0, 500)$. In the square, we simulated food store locations (e.g., features of the built environment) by sampling from an inhomogeneous Poisson point process. The intensity of the inhomogeneous Poisson process was simulated from a log Gaussian process with mean μ_x, marginal variance σ_x^2, and exponential correlation function. In other words, the correlation between two points on the 500×500 grid is given by $\rho(d; \phi)$, where d is the distance between two points and ϕ is the decay parameter, i.e., the rate at which the correlation decays.

We considered three scenarios for the spatial variability of the intensity function: 1) the marginal variance of the intensity function σ_x^2 is set equal to 0; this implies that the intensity is constant over space and store locations are sampled from a homogeneous Poisson point process with a intensity $log(\mu_x)$; 2) $\sigma_x^2 = 1$ and $\phi = 5/3$; this corresponds to a intensity with a spatial correlation that is equal to 0.05 when the distance between two points is equal to 5 units. The resulting sampled locations display a small amount of clustering; and 3) $\sigma_x^2 = 1$ and $\phi = 20/3$; this corresponds to a correlation function that decays to 0.05 at a distance of 20 units, resulting in sampled food stores that display a large amount of clustering. In each case, the mean of the log Gaussian process used to simulate the intensity of the inhomogeneous Poisson process was taken to be equal to 0.15 (see Supplementary Figure D.1).

For each of three built environment settings, we simulated one realization of the built environment, however, given a realization of the built environment, we simulated 1000 datasets with different locations for the health outcomes (e.g., the schools in our motivating application) and different outcome values (e.g., the average BMIz at the various schools).

To simulate school's locations within the $(0, 500) \times (0, 500)$ region, we proceed as follows: we sampled $n \in \{1,000, 6,000\}$ school (x_i, y_i) coordinates from a $Uniform(0, 500)$ distribution, for $i = 1, \ldots, n$. Finally, after counting the number of locations in the built environment around each outcome location, we obtained, for each location i, $X_i(r_{l-1}; r_l)$. We used to generate values of outcome Y_i by sampling from the model: $Y_i = \Sigma_{l=1}^{L} \beta(r_l) X_i(r_{l-1}; r_l) + \epsilon_i$, where $r_0 = 0, r_L = 10, L = 100$ and $\epsilon_i \sim N(0, \tau^2)$. We used two function shapes for $\beta(r)$: 1) a step function given by $\beta(r) = 0.1$ if $r \leq 5$ and 0 otherwise, which results in the true data generating model $Y_i = 0.1X_i(0; 5) + \epsilon_i$ (Figure 3.2a), and 2) a smooth function $\beta(r) = 0.1 f_Z(r)/f_Z(0)$, where $f_Z(r)$ is a normal density function with mean 0 and standard deviation 5/3 (Figure 3.2b). Note that in the traditional models used to study the effect of the built environment on health, the tacit assumption is that the effect of the environment on health can be described by a step function of distance; in other words, the association $\beta(r_{l-1}; r_l)$ is deemed constant up to specified distance r_k but is zero beyond r_k. A step function $\beta(r)$ is likely unrealistic since it is hard to believe that the association abruptly vanishes beyond distance 5, yet this assumption is frequently (implicitly) made in practice. In contrast, the second function used for $\beta(r)$ implies that the association decays smoothly with distance and is near zero by distance 5. We chose the variance τ^2 of the error term so that the model R^2 was equal to either 0.2, 0.5 or 0.8 for the three different built environment schemes. In our motivating example the number of available schools is near 6,000, and the model R^2 was near 0.2 when the DLM was fitted without adjustment of confounders.

Figure 3.2: **True function $\beta(r)$. (a) $\beta(r) = 0.1$ if $r \leq 5$, 0 otherwise. (b) $\beta(r) = 0.1 f_Z(r)/f(0)$, where $f_z(r)$ is a normal density with mean 0 and standard deviation 5/3.**



In fitting DLMs, we chose 100 lags, $L = 100$, with $r_L = 10$. We fitted the model within a Bayesian framework and specified the following prior distributions $\beta_0 \propto 1, \boldsymbol{\alpha} \propto \mathbf{1}, b_1 \sim N(0, \sigma_b^2 I_{L-2}), \sigma_b^2 \sim IG(0.1, 1 \times 10^{-6})$, and $\tau^2 \sim IG(0.1, 1 \times 10^{-6})$. Details on posterior inference and the MCMC algorithm are provided in the Appendix D. For comparison, we also fitted the traditional linear model, $Y_i = \beta_0 + \beta_1 X_i(0; r_k) + \epsilon_i$ which assumes a constant effect of the built environment up to a distance r_k. We used $r_k = 2.5, 5$, and 7.5, respectively, and compared the estimate of β_1 with the estimate of $\bar{\beta}(0; r_k)$ obtained from the DLM for these three distances.

To examine how well DLMs capture true buffer effects at given distance lags, bias, variance, mean squared error (MSE), and coverage rate were calculated at each

$r_l, l = 1, 2, \ldots, L$, using the formulas:

$$Bias(r_l) = \Sigma_{i=1}^{1000}(\hat{\beta}_i(r_{l-1}; r_l) - \beta(r_{l-1}; r_l))/1000,$$
$$Var(r_l) = \Sigma_{i=1}^{1000}\widehat{Var}(\hat{\beta}_i(r_{l-1}; r_l))/1000,$$
$$MSE(r_l) = \Sigma_{i=1}^{1000}(\hat{\beta}_i(r_{l-1}; r_l) - \beta(r_{l-1}; r_l))^2/1000,$$
$$Coverage(r_l) = \Sigma_{i=1}^{1000} I(\hat{\beta}_{i,2.5\%}(r_l) \leq \beta(r_l) \leq \hat{\beta}_{i,97.5\%}(r_l)).$$

To summarize their overall performance and compare DLMs with classical regression models, we calculated the integrated MSE, $IMSE = \frac{r_L}{L}\Sigma_{l=1}^{L} MSE(r_l)$, for both models. In evaluating $IMSE$ for the classical regression models, $\hat{\beta}(r_{l-1}; r_l)$ was set equal to $\hat{\beta}_1$ for $r_l \leq r_k$ and zero otherwise.

When the true DL coefficient function $\beta(r)$ is a step function bias occurs around distance lags where the step happens (Figure 3.3a and 3.3b). Since the fitted DLM assumes that the buffer effect is a continuous function of distance, bias at those lags is expected, and that results in low coverage rates as well. When $\beta(r)$ varies continuously in r (3.3c and 3.3d), much less bias is present, and the bias primarily occurs at the smallest lags because the estimated buffer effects are smoother than the true $\beta(r)$. Some degree of over-smoothing is expected to occur when using random effect variances (vs GCV) to compute smoothing parameters (*Ruppert et al.*, 2003). Also, at the first few lags, there is relatively smaller amount of information since many DL covariates $X_i(r_{l-1}; r_l)$ in the first few lags have many zero values. Hence bias at smallest lags is expected. Additionally, when the degree of clustering in the built environment becomes large, the range of lags at which bias occurs becomes wider and coverage rates tend to be smaller.

For both functions $\beta(r)$, variance of the estimated buffer effects is larger at the first few distance lags due to less information in DL covariates as previously explained. Note also that the variance of the estimated coefficients at both end points (r_l= 0.1 and 10) tends to be larger than for other values of r_l because at the end points the

coefficients are constrained only in one direction. The estimated buffer effects are more variable when the spatial dependence in the intensity function controlling the spatial distribution of the built environment features decays at a slower rate. This can be anticipated because the amount of independent contributions of built environment covariates $X_i(r_{l-1}; r_l)$ is decreased (compare Figure panels 3.3a vs. 3.3b, and 3.3c vs. 3.3d). The MSE is primarily dominated by bias since the variance is fairly constant across a range of distances, except at the endpoints, as mentioned above.

The comparison of estimated average association up to distance r_k, with $r_k = 2.5, 5$, and 7.5, obtained from DLMs and traditional linear models is reported in Table 3.1. The true average association up to distance r_k, $\bar{\beta}(0; r_k)$, is calculated using $\Sigma_{l=1}^{k} \beta(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)/\pi r_k^2$. When locations of food stores are generated from a homogeneous Poisson point process, the estimated associations from the traditional linear models are very close to the true values and their coverage rates are close to 95% (i.e., valid inference) for both functions used for $\beta(r)$. However, if there is clustering of locations in the built environment, the estimated associations from the traditional models are positively biased (away from the null) giving invalid inference unless the model is correctly specified (i.e., when $\beta(r)$ is the step with $r_k = 5$). In particular, when $r_k = 2.5$, a huge amount of bias occurs in the traditional models due to failure in adjusting the effects at longer lags. Note that when negative and positive bias is canceled up to specified distances in the fitted DLMs, bias in estimating the average buffer effect is close to zero (Figure 3.3). In general, compared to the traditional regression models, estimated average buffer effects obtained using DLMs generally performed better having much less bias and better coverage rates except when the fitted traditional models coincide with the true data generating models

Since both the traditional regression models and the DLMs have some degree of bias, we summarize their relative performance in terms of integrated mean squared error (IMSE) up to distance $r_L = 10$ (Table 3.2) (*Ruppert et al.*, 2003). When the

Figure 3.3:

Bias, variance, MSE, and coverage rate at each $r_l, l = 1, 2, \ldots, 100$ for the cases when $\beta(r)$ is: (a) a step function under the built environment without clustering. (b) the step function under the built environment with a large amount of clustering. (c) $\beta(r)$ is the normal *pdf* under the built environment without clustering. (d) $\beta(r)$ is the normal *pdf* under the built environment with a large amount of clustering. Reported results are from a simulation case with $n = 6,000$ and $R^2 = 0.2$.

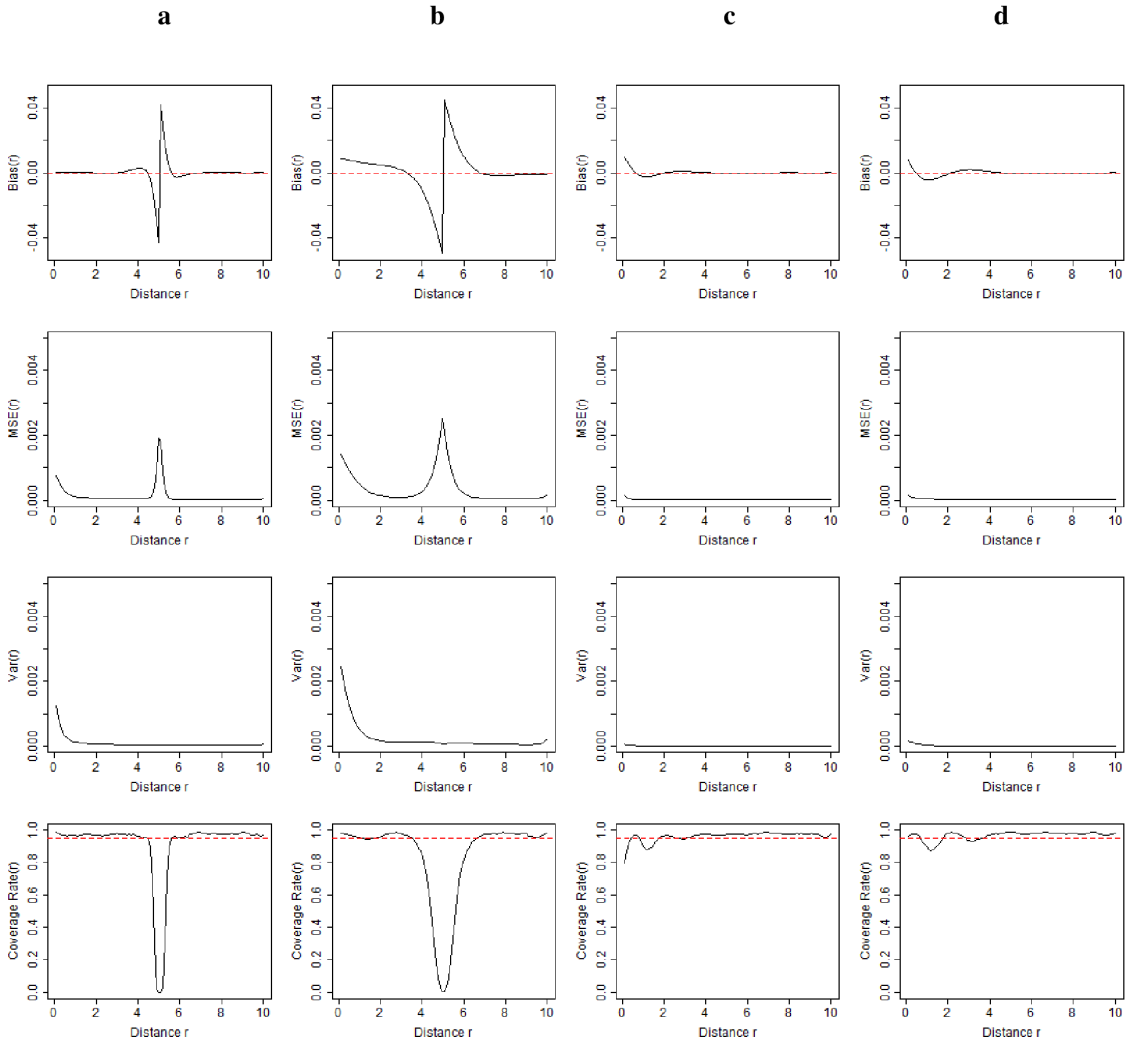


Table 3.1: **Simulation results for the averaged buffer effects up to distance r_k = 2.5, 5, and 7.5 from the traditional model and the fitted DLM. Reported results are from a simulation case with n=6,000 and R^2=0.2.**

			$r_k = 2.5$		$r_k = 5$		$r_k = 7.5$	
$\beta(r)$	Fitted model	Spatial range in the built environment	Est. beta	Coverage rate	Est. beta	Coverage rate	Est. beta	Coverage rate
Step		True $\bar{\beta}(0; r_k)$	0.100	-	0.100	-	0.044	-
	Traditional linear models	Independence	0.103	0.930	0.100	0.949	0.044	0.946
		5	0.241	0.000	0.100	0.939	0.051	0.034
		20	0.304	0.000	0.100	0.953	0.047	0.351
	DLM	Independence	0.100	0.952	0.097	0.790	0.044	0.946
		5	0.103	0.943	0.092	0.500	0.045	0.946
		20	0.105	0.933	0.090	0.554	0.045	0.927
Curve		True $\bar{\beta}(0; r_k)$	0.058	-	0.021	-	0.010	-
	Traditional linear models	Independence	0.059	0.939	0.021	0.947	0.009	0.950
		5	0.074	0.000	0.024	0.064	0.011	0.012
		20	0.078	0.000	0.022	0.464	0.010	0.175
	DLM	Independence	0.058	0.938	0.021	0.950	0.010	0.947
		5	0.057	0.933	0.021	0.947	0.010	0.957
		20	0.057	0.933	0.021	0.950	0.009	0.948

Table 3.2: **Integrated MSE from fitted traditional linear models with distance lag r_k = 2.5, 5, and 7.5 and from fitted DLMs with a maximum distance r_L = 10. Reported results are from a simulation case with $n = 6,000$ and R^2=0.2.**

		Traditional linear model (r_k=2.5)	Traditional linear model (r_k=5)	Traditional linear model (r_k=7.5)	DLM
$\beta(r)$	Spatial range in the built environment	IMSE*	IMSE*	IMSE*	IMSE*
Step	Independence	2.512	0.004	2.055	0.125
	5	7.500	0.004	1.861	0.269
	20	12.883	0.003	1.945	0.360
Curve	Independence	0.205	0.790	1.113	0.006
	5	0.169	0.743	1.060	0.010
	20	0.179	0.767	1.083	0.013

Note: IMSE* is multiplied by 100.

true form of the $\beta(r)$ function is the step function, the IMSE was minimum for the traditional regression models using the a-priori distance lag $r_k = 5$, which is not surprising since the estimated model is the data generating model. However, when $\beta(r)$ decays with distance r, the DLMs consistently yield the smallest IMSE.

To conserve space and avoid redundancy, here we only reported results for the simulation setting with $n = 6000$ and $R^2 = 0.2$, since this scenario corresponds to the data in our motivating example. For the smaller sample sizes (tables and figures shown in the Appendix), bias and coverage rates of the DLM estimates deteriorate, and the strong confounding bias in the traditional regression models persists. The bias in the DLM is largely attenuated when the model R^2 increases, but this does not happen for the traditional regression models.

To further examine assumptions used by the fitted DLMs we conducted additional simulations: 1) we specified different numbers of lags, i.e., L=25, 50, 200, to define ring-shaped areas that differ from the ones (L=100) used in the data generating model; and 2) we assumed different maximum distance r_L=3, 20. As expected,

using a smaller numbers of lags in DLMs (L=25), resulted in smoother estimated DL coefficients because the DL coefficients are estimated in wider ring shaped area and thus become coarser. A larger number of lags (L=200) yielded similar results as L=100. When the maximum distance was misspecified and r_L=3, we observed bias in the DL coefficients when there is clustering of locations in the built environment. However, the amount of bias in estimates of the average buffer effect at r_k=2.5 was less than that from traditional regression models. Results were consistent to those with r_L=10 when the maximum lag distance used to fit the DLMs was equal to 20.

3.4 Data Example

We analyzed the FitnessGram data for 5^{th} and 7^{th} grade children who attended public schools in California in 2010. The FitnessGram data is publicly available from the California Department of Education (CDE) and includes measures of children's weight and height as well as other individual characteristics (e.g., grade, age, gender, race/ethnicity). We averaged children's BMI z-score (BMIz) within a school and used it as the outcome. BMIz is an age and gender-adjusted BMI measure because the meaning of BMI is not the same across growing children of different age and sex (*CDC*, 2005). Similarly, we averaged the characteristics of the students whose BMI was recorded and we used these averages as covariates: they are percentage of 7^{th} graders, percentage of female students, percentage of Hispanic children and percentage of other ethnicities (Asian, African American, and Filipino combined).

The location of convenience stores (CS) in California was purchased from a commercial source (National Establishment Time Series from Wall and Associates). Geocodes for schools and food stores were cross-referenced to obtain the number of CS between two radii r_{l-1} and $r_l, l = 1, \ldots, 100$, from a school with a maximum lag distance of $r_{100} = 7$ miles. We also obtained other school characteristics from the CDE, namely, total student enrollment in the school, percentage of children in the school that par-

ticipated in the California school free or reduced meal program, and percentage of adults with a bachelor's degree or higher residing in the schools' census tract; the latter was obtained from the 2000 US Census.

The total number of schools in the dataset was 5,903, while the mean (SD) for the number of CS within 1/4, 1/2, and 3/4 miles from schools is, respectively, 0.16 (0.47), 0.68 (1.06), and 1.49 (1.8). The overall mean (SD) BMIz for schools in the dataset is 0.75 (1.09).

We started our analysis fitting traditional, weighed linear models $Y_i = \beta_0 + \beta_1 X_i(0; r_k) + \epsilon_i$, where Y_i is the average BMIz in school i, $X_i(0; r_k)$ is the number of CS up to distance r_k from a school i, and $\epsilon_i \sim N(0, \tau^2/n_i)$ with n_i number of children in a school i, where r_k was respectively equal to 1/4, 1/2, or 3/4 miles from schools. We also fitted the weighted DLM, following the approach discussed in Section 3.2.4, with weights set equal to the number of children in each school whose BMI was measured.

We fitted both crude models without any adjustment as well as models adjusting only for the average student characteristics and models adjusting for both the average student characteristics and the school characteristics. School neighborhood characteristics (school's neighborhood socioeconomic position) can act as confounders or mediators (*Chaix et al.*, 2010; *Diez Roux*, 2004) since it is uncertain if low neighborhood socioeconomic conditions were driven by poor-quality food environments or, vice versa, if poor-quality food environments are caused by low neighborhood socioeconomic position. Hence, we included both results.

The estimated average effects of CS up to 1/4, 1/2, and 3/4 miles on children's BMIz from the weighted linear regression models and the DLMs are summarized in Table 3.3. In the crude analysis, the estimated average associations between the number of CS up to 1/4, 1/2, or 3/4 miles were all significant for both the weighted regression models and DLMs. Interpreting the results from the DLM, for instance,

Table 3.3: **The estimated buffer effects of CS up to buffer sizes 1/4, 1/2, 3/4 miles from the traditional linear models and DLMs.**

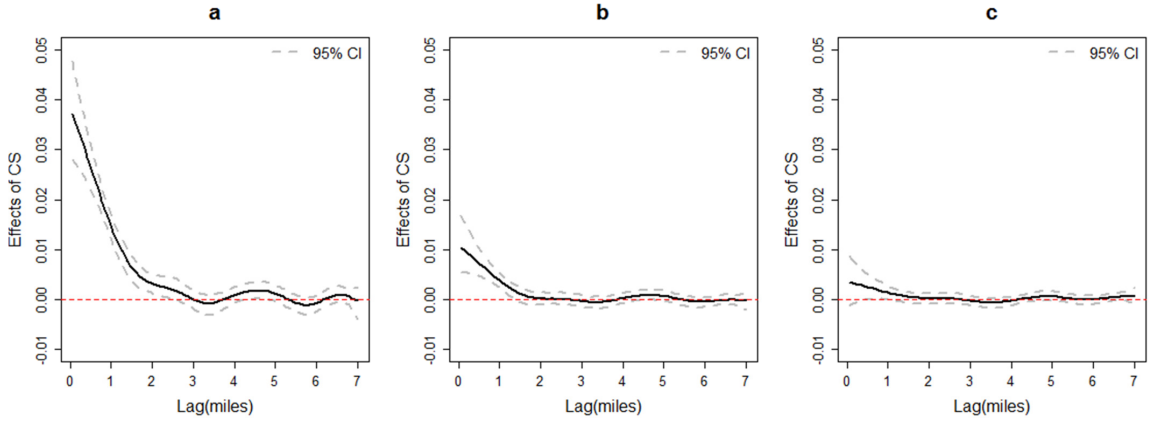
Model	A specified distance	Traditional linear models			DLMs		
		beta	LCI	UCI	beta	LCI	UCI
Crude	CS(1/4 mile)	0.092	0.076	0.109	0.035	0.026	0.044
	CS(1/2 mile)	0.072	0.064	0.079	0.030	0.024	0.036
	CS(3/4 mile)	0.051	0.047	0.056	0.026	0.021	0.030
Adjusted school's student participant characteristics	CS(1/4 mile)	0.023	0.012	0.034	0.009	0.005	0.014
	CS(1/2 mile)	0.016	0.011	0.021	0.008	0.005	0.012
	CS(3/4 mile)	0.011	0.008	0.014	0.007	0.004	0.010
Adjusted school's student participant characteristics + school characteristics	CS(1/4 mile)	0.011	0.001	0.021	0.003	0.000	0.007
	CS(1/2 mile)	0.007	0.002	0.011	0.003	0.000	0.006
	CS(3/4 mile)	0.005	0.002	0.008	0.002	0.000	0.005

we expect the mean children's BMIz to be 0.035 (95% CI: 0.026, 0.044) higher for each additional CS within 1/4 mile from schools. Adjusting for average student characteristics or for both average student and school's characteristics, attenuated all coefficients. Overall, the coefficients from the weighted linear regression model tend to be larger (approximately 2 times larger), although this is likely due to over-estimation as we observed in the simulations. It is probably due to the presence of spatial correlation in the built environment.

Figure 3.4 (a)-(c) present the estimated coefficients for CS within 7 miles from schools as estimated from the fitted DLMs. The crude DL coefficients were significant up to a distance of approximately 2.5 miles and within 5 miles from schools; additionally, as it might be expected, they were highest for distances that are within walking distance. After adjusting for the average student characteristics, the DL coefficients were highly attenuated but they remained significant within shorter distances from the schools. After further adjustment for school characteristics, effects became even more attenuated and were almost not significant except for within short walking distance.

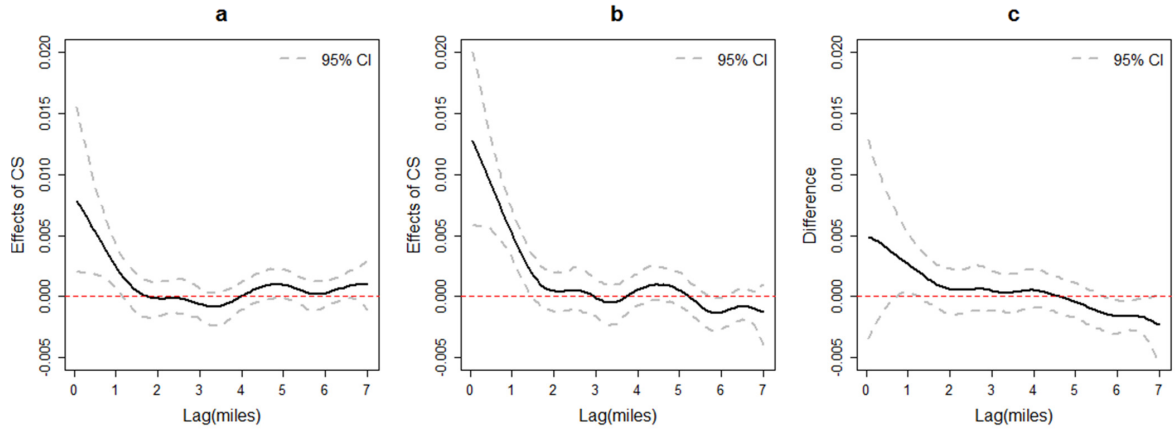
Further, we investigated whether the associations were different for 5^{th} grade chil-

Figure 3.4: **The estimated DL coefficients of CS up to 7 miles from schools (a) without the adjustment of confounders , (b) with the adjustment of school's student participant characteristics, and (c) with the adjustment of both school's student participant characteristics and school's characteristics.**



dren versus 7^{th} grade children. Our hypothesis is that this would be the case since the 7^{th} graders might have different behaviors or more ability to walk further. To assess this, we used the approach discussed in Section 3.2.4 and included in the model the percentage of 7^{th} grade children in the school as an interacting covariate. Although the interaction was not significant in the crude models, it was significant in the models where we adjusted for the average students' characteristics. Figure 3.5 shows the estimated DL coefficients relative to distances up to 7 miles for schools where the participants are only 5^{th} grade children (Figure 3.5a), 7^{th} grade children (Figure 3.5b), and the difference of between the two (Figure 3.5c). Estimated coefficients for short walking distances from schools where participants are only 7^{th} grade children were greater than those for schools whose participants are only in 5^{th} grade. Adjusting for both the average student characteristics and the school characteristics attenuated the effect sizes while keeping a pattern similar to the one displayed in Figure 3.5.

Figure 3.5: **The estimated DL coefficients up to 7 miles from schools where participants are only (a) 5^{th} grade children or (b) 7^{th} grade children, and (c) the difference of buffer effects for schools between only 5^{th} grade participants and only 7^{th} grade participants.**



3.5 Discussion

We have proposed a distributed lag model (DLM) to examine associations between built environment factors and health. This flexible model allows us to examine how associations between features of the built environment and health are distributed up to a maximum distance from sample locations. The DLM approach is based on constructing built environment measures within L ring-shaped regions (DL covariates) around sample locations, rather than buffers. Based on distributed DL coefficients, we are able to calculate average buffer effects up to a chosen distance $r_k, k \leq L$, without assuming constant effects up to that distance; traditionally, the most common approach used in epidemiology, ecology, and transport geography impose such an assumption.

The maximum distance in DLMs implies that we expect no further association between the outcome and the built environment factors beyond that maximum distance. Violation of this assumption might cause bias in estimation since the DL coefficients would be confounded by associations with features beyond the maximum distance when spatial correlation exists in the built environment. While the tradi-

tional approaches require the user to speculate the distance where effects may be present, a DLM has a less stringent requirement by specifying the maximum distance beyond which effects are zero and simultaneously allow us to examine if these effects are indeed vanishing with distance.

We have compared the performance of the traditional models with that of DLMs through a simulation study that comprised various scenarios for degree of clustering in the built environment as well as different functional forms for the DL coefficients. Our simulations confirmed that the performance of the DLMs is superior to that of traditional models in terms of bias and coverage rates of true effects. Our results expand on the work by *Spielman and Yoo* (2009), who used simulation settings slightly different from ours, and discussed only results relative to bias. We have also examined how well DLMs capture true DL coefficients for various distances from sample locations. We have found that the estimated coefficients relative to the first few lags had more uncertainty compared to later lags due to the fact that they could exploit relatively less information, i.e., more zero values in the first few lags. We also observed inflated variance of estimated DL coefficients at both end points since constraints are imposed only in one direction. Exploring whether other types of constraints for the DL effects attenuate this limitation is a needed next step; possible constrains include variations of those introduced by *Welty et al.* (2009). Alternatively, one could assume that the DL coefficients follow a parametric function as was initially done in DL models. Another possible direction is to use kernel averaged predictor models (*Heaton and Gelfand*, 2011). With additional constrains, potentially derived from substantive knowledge, efficiency may be gained by imposing assumptions such as coefficients being zero after a given distance. Our proposed model does not make strong assumptions on the form of the DL coefficients except selecting the maximum distance.

In our application, we have examined the effect of the availability of convenience

stores (CS) on child's BMI z-score (BMIz) using a surveillance dataset for 5^{th} and 7^{th} grade children in the 2010 FitnessGram. Comparing results between the traditional models and the DLMs, we found that estimates from the traditional models were usually higher than those from DLMs due to the overestimation that results from the presence of spatial correlation (i.e., clustering of CS) in the built environment. Applying the proposed DLM, we have been able to investigate whether there is a difference in the DL coefficients by types of student participant in the school, and have found that the DL coefficients are higher when the percentage of 7^{th} grade children in schools is higher. Additionally, DL coefficients were significant at slightly longer lags compared to schools where there was a higher percentage of 5^{th} grade children participants. DLMs can help identify if the distance within which built environment factors affects health varies by subject characteristics.

Although distributed lag models have a long history, this is the first application of DLMs to study the associations between health and the built environment. This innovative application of DLMs can help shed light on the relevant distances within which the built environment may associate with health.

CHAPTER IV

Hierarchical Distributed Lag Models: heterogeneity in associations between the Built Environment and Health

4.1 Introduction

Over the last several years, research to identify contributors to the childhood obesity epidemic has dramatically increased, owing to the tremendous short and long term costs associated with this condition (*Daniels*, 2006; *Reilly and Kelly*, 2011; *Reilly et al.*, 2003). Beyond individual level predictors of obesity, areal level factors, particularly specific features of the built environment may also contribute to poor diet and inactivity and thus influence body weight. As children spend a large amount of time in schools, one feature of interest is the food environment near schools. The food environment can be conceived from a broad point of view as encompassing food advertisements, to presence of certain food outlets, to more detailed information such as the quality and quantity of available foods. Research has found that convenience stores provide ready access to cheap, low nutrient, high calorie food items, and their presence near schools has been linked to child obesity (*Kipke et al.*, 2007; *Rahman et al.*, 2011; *Sallis and Glanz*, 2006; *Singh et al.*, 2010).

However, the link between convenience store availability near schools and chil-

dren's obesity has been inconsistent potentially due to several reasons. One possible reason is between-study differences in the geographic scale used to measure features of the food environment (*Black and Macinko*, 2008; *Schaefer-McDaniel et al.*, 2010). Some studies use counts of convenience stores within a circular area, or "buffer", centered at locations of interest with a pre-specified radius, while others use zip-code, census tract, or county-level counts. The question of how to choose the geographic scale to construct such measures is widely known as the modifiable areal unit problem (MAUP) (*Fotheringham and Wong*, 1991; *Openshaw*, 1996), and incorrect selection of the appropriate geographic scale can lead to severe bias in the association of interest (*Baek et al.*, 2014; *Spielman and Yoo*, 2009). Additionally, it is possible that individuals' true geographical context varies across study locations, and this is often referred to as the uncertain geographic context problem (UGCop) (*Kwan*, 2012). For instance, children's activity spaces may vary according to the degree of street connectivity or availability of sidewalks. Not accounting for differences in activity spaces translates to measurement error in the covariate of interest and thus results in bias and incorrect inference. Alternatively, these food environment-obesity associations may truly vary across places, potentially due to differences in unobserved properties of the food environment, such as food quality (*N.P.*, 2014), food policy and/or obesity prevention policies. Many studies have used commercial databases where food establishments can be classified into several different types, such as 'convenience stores' and 'fast food restaurants', but obtaining data on the actual nutritional quality of food items sold at these food outlets is not always possible in part because of the tremendous cost to obtain such data.

The goal of this study is to systematically examine variations in the association between the food environment and children's body weight across assembly districts in California. We use information from two large scale databases: (1) the Fitnessgram database, a surveillance database containing body weight information of essentially

all 7^{th} grade children who attended California public schools during the period 2001-2010, and (2) the National Establishment Time Series from Walls and Associates; this latter dataset contains yearly information on the availability of food outlets in the entire state. There are several reasons for choosing assembly districts as areal units in studying variation in the built environment associations with health. First, previous reports have found that there are large variations in childhood obesity across assembly districts (*Drewnowski et al.*, 2009); additionally, regional differences in the food environment quality across California have been previously documented (*N.P.*, 2014). Thus, it is possible that the associations between body weight and features of the built environment measured using large scale databases may vary due to unmeasured information, such as food quality. Second, assembly districts may be an important level of aggregation because they are relevant for policy making; they are politically active units with representation in the state legislature and have the potential to stimulate regulation of food environments around schools (*N.P.*, 2013). Third, research that explicitly considers assembly districts as units of analyses can help inform future population-wide obesity prevention interventions at this level.

To achieve our goal, we propose a hierarchical distributed lag model (HDLM) extending a DLM recently applied in the built environment research (*Baek et al.*, 2014). DLMs are useful for (a) examining associations between features of the built environment and health over distance around locations of interest (e.g., schools) and (b) calculating average buffer effects up to a chosen distance (e.g., a 1 mile buffer) more accurately than with traditional linear models. With the proposed hierarchical extension of the DLM, we allow assembly districts to have their own geographic scale by modeling the magnitude and shape of the DL coefficients as random effects across assembly districts.

Baek et al. (2014) showed that when there are few food outlets near schools, the variance of the DL coefficients will increase. In the present application, DL coefficients

from more rural assembly districts may have higher variance due to fewer food outlets near schools compared to urban-assembly districts. In the proposed model, we expect that the estimated assembly-district specific DL coefficients will be shrunk to the overall DL coefficients; thus, districts with less information in the DL covariates can borrow strength from other assembly districts. By doing so, HDLMs enable us to control the variance in the DL coefficients relative to assembly districts with less data. Although differences in DL coefficients according to pre-specified covariates can be examined (*Baek et al.*, 2014), examining the variation via hierarchical models allows us to take an agnostic approach to examine and quantify any potential variation in associations, and this in turn will enable researchers to generate hypotheses about the sources of variation in the DL coefficients.

HDLMs have previously been implemented in the air pollution literature. Researchers investigated the overall temporal effects of air pollutant exposures on health outcomes through combing information across areas (or groups) and some others further examined heterogeneity of the effects across areas (*Berhane and Thomas*, 2002; *Dominici et al.*, 2000; *Huang et al.*, 2005; *Madden and Paul*, 2010; *Peng et al.*, 2009; *Rondeau et al.*, 2005; *Zhao et al.*, 2014). Several examples of HDLM approaches implemented within a Bayesian framework are found in *Dominici et al.* (2000), *Huang et al.* (2005), *Peng et al.* (2009) in the air pollution literature. However, in all these examples HDLMs are fitted using a two-stage approach, reducing computational costs. However, using a two-stage approach in our application would fail to control the inflation of variance in the estimates due to sparse covariate information in some areas. In this paper, we jointly estimate parameters of HDLMs in one-step procedure using a Bayesian framework.

4.2 Data Sources

We examined FitnessGram data for 7^{th} grade children who attended public schools in California in 2001-2010. The FitnessGram dataset is publicly available from California Department of Education (CDE) and includes measures of children's weight and height and other individual characteristics (e.g., grade, age, gender, race/ethnicity). We averaged children's BMI z-score (BMIz) (CDC growth charts, 2005) within a school and used it as the outcome. Similarly, we averaged characteristics of students participating in data collection to use as covariates: percentage of female students, percentage of Hispanic children and percentage of other ethnicities (Asian, African American, and Filipino combined).

Location of convenience stores (CS) in California was purchased from a commercial source (National Establishment Time Series from Wall and Associates). Geocodes for schools and food stores were cross-referenced to obtain the number of CS between two radii r_{l-1} and r_l, $l = 1, \ldots, 100$, from a school with a maximum lag distance of $r_{100} = 7$ miles.

We also obtained other school characteristics from the CDE, namely, total student enrollment in the school, percentage of children in the school that participated in the California school free or reduced meal program, and percentage of adults with a bachelor's degree or higher residing in the schools' census tracts, the latter obtained from the 2000 US Census. We used the assembly district boundaries set in 2001 rather than in 2011, since FitnessGram data were from the period 2001-2010. We obtained the 2001 assembly districts' shapefile at http://statewidedatabase.org/geography.html.

4.3 Exploratory Analysis

We performed several exploratory analyses to guide our model building strategy. First, we examined the spatial pattern of the assembly district means of children's

BMIz, overall and by year, to determine if spatially correlated assembly-district intercepts would be needed, and if the spatial pattern of the district means changed across time. While there was evidence of spatial correlation in the means (Moran's I p-value < 0.001), the spatial pattern was similar over time (Appendix E Figure E.1), meaning that modeling spatially correlated district means might be needed, but modeling space-time interaction of district specific means is not necessary. Second, we explored the heterogeneity and smoothness of DL effects across assembly districts by fitting the non-hierarchical DLM in the data stratified by assembly districts. Although it is possible that each assembly district requires its own smoothness parameter (Appendix E Figure E.2a), we also considered grouping assembly districts into two groups (Appendix E Figure E.2b and E.2c) since in the exploratory analyses the estimated shapes of DL coefficients were either nearly linear or have a similar degree of smoothness. Finally, we examined spatial correlation in 1/4, 1/2, 3/4 mile buffer effects across assembly districts to decide whether random DL coefficients of assembly districts further need to be spatially structured, from which we concluded that modeling spatial correlation of the district specific DL coefficients would not be needed (p-values of Moran's I were 0.67, 0.66, 0.63 at 1/4, 1/2, 3/4 miles, respectively) (Appendix E Figure E.3).

4.4 Hierarchical Distributed Lag Models (HDLMs)

Let Y_{ijt} be the average BMIz among n_{ijt} children attending school i in assembly district j at time t (schools are the unit of observation), and let $X_{ijt}(r_{l-1}; r_l), l = 1, 2, \ldots, L$, be the number of convenience stores (CS) between two radii r_{l-1} and r_l around school i in assembly district j at time t. Time t is years since 2001, and ranges from 0 to 9. The distance r_L is the maximum distance around schools (the maximum buffer size with a radius equal to r_L) after which we assume no further association between the measured feature and the outcome. The total number of lags, L, can be

chosen large enough so that the DL effects have a smoother effect as a function of distance, in our case L=100.

We build the model in a hierarchical fashion to achieve an increased flexibility in the way the DL coefficients are modeled. Our baseline HDLM (model 1) is assumed to have constant DL effects across assembly districts

$$Y_{ijt} = \beta_0 + \sum_{l=1}^{L} \beta(r_{l-1}; r_l) X_{ijt}(r_{l-1}; r_l) + \mathbf{Z}_{ijt}\boldsymbol{\gamma} + \mathbf{T}_{ij}\boldsymbol{s}_{ij} + \mathbf{T}_{ij}\boldsymbol{\eta}_j + \epsilon_{ijt}, \qquad (4.1)$$

where β_0 is the overall mean BMIz, $\beta(r_{l-1}; r_l)$ is a DL effect of the environment feature measured between two radii r_{l-1} and r_l around schools, $\mathbf{Z}_{ijt}$ are covariates relative to the fixed effects part of the model and include percentage of female students, percentage of Hispanics, percentage of other ethnicities, and time modeled as a linear spline with a knot at year 2005 ($t = 4$). Finally, ϵ_{ijt} represents a residual error assumed to follow a mean-zero normal distribution with variance τ^2. The change of slope for time is included because California adopted food and beverage policies in 2004 to improve public school food environment which was shown to have an effect on obesity rates (*Sanchez-Vaznaugh et al.*, 2010). A knot at year 2005 instead of year 2004 is used since we expect some latency period of the food policy and a knot at year 2005 actually provides better model fit. Moreover, a covariates matrix, $\mathbf{T}_{ij}$, is a subset of $\mathbf{Z}_{ijt}$ and includes an intercept, time, and spline time at year 2005 with corresponding school and district level random effects, $\boldsymbol{s}_{ij}$ and $\boldsymbol{\eta}_j$, assumed normally distributed. With the school and district level random effects, we account for unobserved covariates that may affect BMIz or modify time trends at both school and district levels. Including random spline times of schools and assembly districts may yield rather complex models, but our preliminary analysis showed that including all those terms yielded better fit. Because there is large variability in the number of children per school, we adapted the proposed models to include weights set equal to

the number of children who participated within schools. (*Baek et al.*, 2014).

We constrain the coefficients $\beta(r_{l-1}; r_l)$ to vary as a smooth function of distance $r_l, l = 1, 2, \ldots, L$, from schools using splines (*Hastie and Tibshirani*, 1990; *Zanobetti et al.*, 2000). Constraining the coefficients ensures coefficients corresponding to adjacent ring-shaped areas to be similar, since we would not typically expect associations to change abruptly across distance, and also controls possible numerical problems that may arise when many schools have zero CS between given radii r_{l-1} and r_l. We used cubic smoothing splines to constrain the association coefficients $\beta(r_{l-1}; r_l)$ implemented using a radial basis function,

$$\beta(r_{l-1}; r_l) = \alpha_0 + \alpha_1 r_l + \sum_{k=1}^{L} \tilde{\alpha}_k |r_l - r_k|^3, \tag{4.2}$$

where α_0 is the global intercept of the lag effects, and α_1 is the global average change rate of lag associations over distance while the coefficients $\tilde{\alpha}_k$ are penalized to achieve smoothness of the global DL coefficients (see Appendix E).

To allow variation in the DL effects across assembly districts, we include random DL coefficients in the equation (4.1),

$$Y_{ijt} = \text{model 1} + \sum_{l=1}^{L} b_j(r_{l-1}; r_l) X_{ijt}(r_{l-1}; r_l), \tag{4.3}$$

where $b_j(r_{l-1}; r_l) = \alpha_{0j} + \alpha_{1j} r_l + \sum_{k=1}^{L} \tilde{\alpha}_{jk} |r_l - r_k|^3$ is a deviation from the global DL coefficients between radii r_{l-1} and r_l specific to assembly district j.

Based on our exploratory analysis, several variants of the DL model with random DL effects shown in (4.3) were considered (see Table 4.1). Models 2 and 3 assume that there is variability in the DL coefficients at the assembly district level (e.g., variation exists in intercepts and slopes of lag effects), but the same amount of curvature in the random DL coefficients is sufficient to capture the variation. Since we centered distance lags at their mean ($\bar{r} = 3.54$) for numerical stability, random intercepts and

Table 4.1: **Several variants of the HDLM in (4.3) having various random DL effects distributional assumptions.**

Model	$b_j(r_{l-1}; r_l)$
2	$\begin{bmatrix}\alpha_{0j}\\ \alpha_{1j}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_2^2 & 0\\ 0 & \sigma_3^2\end{bmatrix}\right)$ and $\tilde{\alpha}_{jk} = 0$.
3	$\begin{bmatrix}\alpha_{0j}\\ \alpha_{1j}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \mathbf{\Sigma}_\alpha\right)$, where $\mathbf{\Sigma}_\alpha$ is positive definite, and $\tilde{\alpha}_{jk} = 0$.
4	$\begin{bmatrix}\alpha_{0j}\\ \alpha_{1j}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \mathbf{\Sigma}_\alpha\right)$, where $\mathbf{\Sigma}_\alpha$ is positive definite, and $\tilde{\alpha}_{jk} \sim N(0, \sigma_v^2)$.
5	$\begin{bmatrix}\alpha_{0j}\\ \alpha_{1j}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \mathbf{\Sigma}_\alpha\right)$, where $\mathbf{\Sigma}_\alpha$ is positive definite, and $\tilde{\alpha}_{jkg} \sim N\left(0, \sigma_{vg}^2\right)$, $g = 1, 2$.

slopes of the DL coefficients may be no longer independent as mentioned by *Durbán et al.* (2005). For that reason, we compared between model 2 and model 3. Model 4 further assumes that the smoothness in the DL coefficients across assembly districts is different from that of the overall DL effects. Model 5 is constructed to allow for more flexibility in the smoothness of DL coefficients which is categorized into two groups suggested by the exploratory analysis. Note that the models 1-5 are constructed in a nested fashion. We compared models based on the widely used deviance information criterion (DIC) which trades off model fit and complexity (*Spiegelhalter et al.*, 2002).

Given the final chosen model, we further examined whether a spatially structured random intercept of assembly districts improves model fit, as we speculated in the exploratory analysis that spatial autocorrelation of district-specific means may be needed. The spatially structured random intercept is modeled by a conditional autoregressive (CAR) prior distribution, $\eta_j \sim N(\sum_{j\sim j'} \frac{w_{jj'}\eta_{j'}}{w_{j+}}, \frac{\tau_\eta^2}{w_{j+}})$, where $j \sim j'$ denotes assembly district j is a neighbor of assembly district j' defined as sharing any boundary between two districts, $w_{jj'} = 1$ if $j \sim j'$ and 0, otherwise, and w_{j+} is the total number of neighbors for area j (*Besag et al.*, 1991; *Clayton and Kaldor*, 1987).

Similarly as in Chapter 3, we used the selected model to estimate the average difference in children's BMI z-score per one additional CS within a buffer area of a 1/2 mile radius in each assembly district j, $\bar{\beta}_j(0; 1/2)$. The 1/2 mile distance is widely used

in built environment literature concerning children, therefore enhancing the comparability of our results to other work. For a given distance r_k from schools, $\bar{\beta}_j(0; r_k) = \sum_{l=1}^{k} \beta_j(r_{l-1}; r_l)\pi(r_l^2 - r_{l-1}^2)/\pi r_k^2$, where $\beta_j(r_{l-1}; r_l) = \beta(r_{l-1}; r_l) + b_j(r_{l-1}; r_l)$ are DL coefficients for an assembly district j between radii r_{l-1} and r_l.

We took a Bayesian approach for estimation to make inference for all model parameters including the district-specific random effects. This approach incorporates uncertainty of dispersion parameters of random effects and smoothness parameters of DL coefficients. Posterior samples of district-specific DL effects were transformed to easily calculate district-specific buffer effects up to 1/2 mile from schools and respective 95% credible intervals. The sampling approach we use is based on shrinkage slice sampling methods proposed by *Agarwal and Gelfand* (2005) which are an extension of *Neal* (2003). Compared to Metropolis-Hasting algorithms shrinkage slice sampling does not require controlling the acceptance rate of proposed samples from prior distributions, and thus it achieves faster convergence of posterior samples and enhances computational efficiency (see Appendix E for more details).

4.5 Results

Locations of schools, assembly district-specific BMIz and concentration of convenience stores are given in Figure 4.1. Schools are densely located in metropolitan areas as expected (Figure 4.1a). Geographically larger assembly districts tended to have more schools, while assembly districts in large metropolitan areas had fewer schools with a larger number of children per school. In California, the overall mean (SD) BMIz was 0.71 (1.07), and the assembly district-specific BMI means ranged from 0.32 to 0.98 (Figure 4.1b).

The average number of convenience stores within 1/2 mile from schools over the study time period was 0.62 (SD=0.97) and the assembly district-specific means ranged from 0.09 to 1.96 (Figure 4.1c).

Table 4.2: **Descriptive statistics of children's BMIz and the number of CS within 1/4, 1/2, and 3/4 miles from schools in FitnessGram 2001-2010 for 7^{th} grade children.**

					Number of Convenience stores within…					
	N of	N of	BMIz		¼ mile		½ mile		¾ mile	
Year	children	schools	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Overall years	3,564,709	2,190*	0.71	(1.07)	0.16	(0.44)	0.62	(0.97)	1.34	(1.67)
2001	283,227	1,238	0.66	(1.06)	0.13	(0.37)	0.45	(0.78)	0.97	(1.26)
2002	313,745	1,347	0.67	(1.07)	0.13	(0.39)	0.49	(0.83)	1.07	(1.38)
2003	350,323	1,454	0.70	(1.07)	0.13	(0.41)	0.53	(0.88)	1.16	(1.49)
2004	363,187	1,513	0.71	(1.07)	0.15	(0.44)	0.62	(0.96)	1.32	(1.60)
2005	370,634	1,589	0.73	(1.07)	0.15	(0.43)	0.62	(0.97)	1.34	(1.62)
2006	376,764	1,649	0.72	(1.08)	0.16	(0.45)	0.67	(1.02)	1.44	(1.75)
2007	379,263	1,719	0.72	(1.08)	0.17	(0.47)	0.68	(1.02)	1.45	(1.75)
2008	383,024	1,764	0.72	(1.07)	0.18	(0.48)	0.72	(1.07)	1.50	(1.86)
2009	373,017	1,711	0.71	(1.07)	0.18	(0.49)	0.72	(1.06)	1.53	(1.87)
2010	371,525	1,715	0.71	(1.07)	0.16	(0.44)	0.64	(0.97)	1.40	(1.68)

* indicates the number of unique schools in data.

Table 4.2 shows descriptive statistics for the number of 7^{th} grade children, number of schools, children's BMIz and the number of CS within 1/4, 1/2, and 3/4 miles from schools, overall and by year. As shown in Table 4.2, the mean of children's BMIz increases up to 2005 and becomes stable after 2005 due possibly to food and beverage policies adopted in 2004 (*Sanchez-Vaznaugh et al.*, 2010). The number of CS within 1/4, 1/2, or 3/4 mile buffers also increases up to 2004 and becomes stable after that time.

Table 4.3 shows DIC values for the models considered in Table 4.1. Including random DL effects of assembly districts further improves model fit (model 1 vs. model 2), and a positive definite covariance structure of the random coefficients provides a slightly better model (model 2 vs. model 3). However, models allowing more flexibility in the degree of smoothness of random DL coefficients (model 4 and model 5) resulted in worse DIC because improvement in model fit did not offset the increased model complexity. Model 3 had the lowest DIC and was thus selected. The model further including a spatially structured random intercept of assembly districts did not improve

Figure 4.1: **Locations of unique schools (black dots) and number of schools within assembly districts; (b) assembly district mean of BMIz; (c) assembly district mean number of CS within 1/2 mile from schools across CA, LA and SF metropolitan areas. Data Sources: 2001-2010 Fitnessgram data for 7^{th} grade children, CDE; National Establishments Time Series database.**

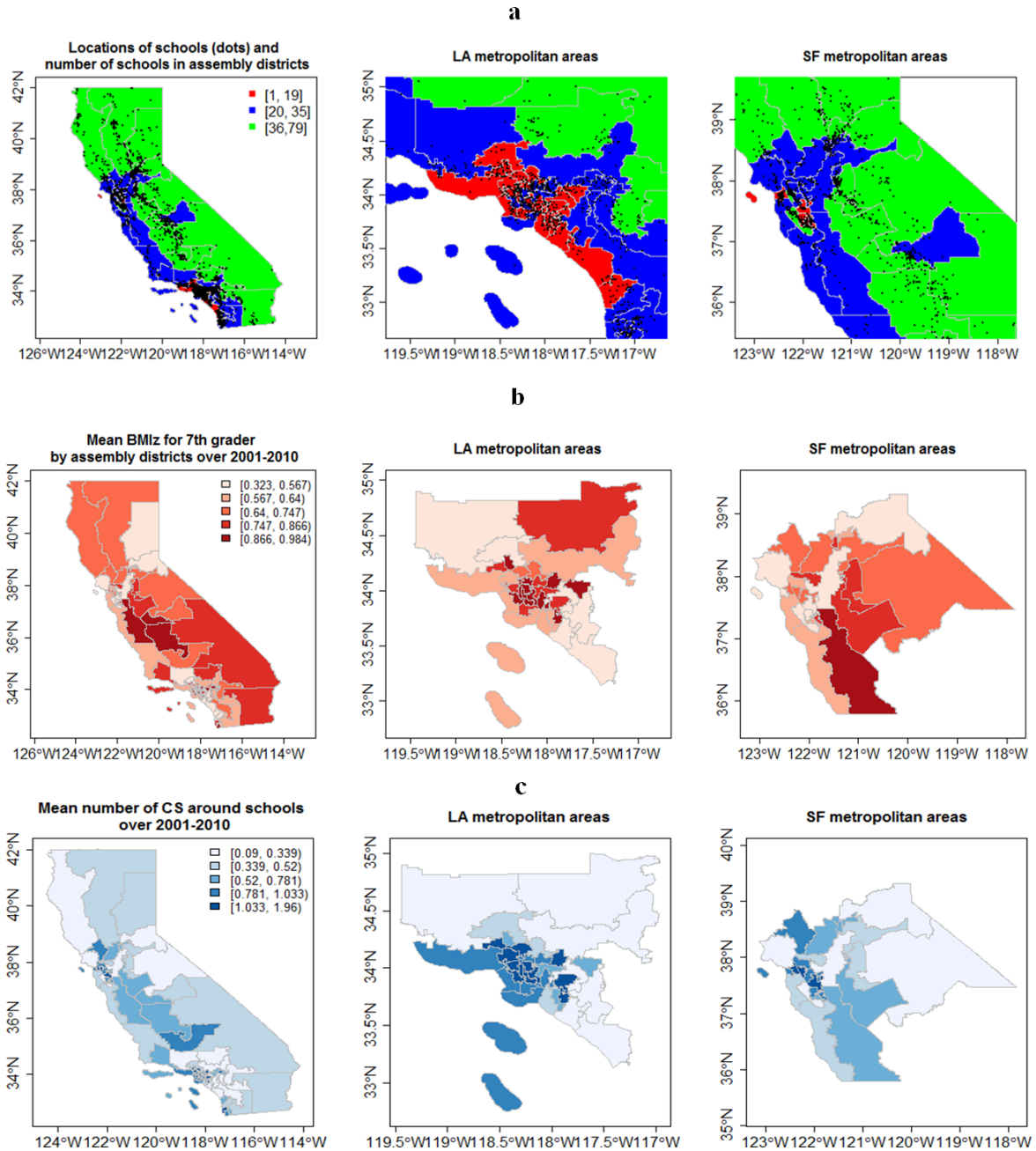


Table 4.3: **Deviance information criterion (DIC) for model selection.**

Model	DIC
Model 1	66,555
Model 2	66,548
Model 3	66,543
Model 4	66,620
Model 5	66,646

model fit (DIC = 66,659), probably because the features of the built environment, which are spatially correlated, explain the spatial pattern of the mean BMIz seen in our exploratory analysis (Figure 4.1 b-c).

Based on the selected Model 3, we estimated the overall DL effects of the measured built environment, the random DL effects of assembly districts, and 1/2 mile buffer effects of assembly districts (Figure 4.2). The estimated overall DL effects (Figure 4.2a) had wide credible intervals at the first few lags due to sparser built environment information at shorter distances (i.e., many schools have zero CS within the first few lags). The overall DL effects became null at around 1.8 miles. Assembly district specific DL coefficients (Figure 4.2b) had the same smoothness as the overall DL coefficients, but with different intercept and slope of the district-specific DL coefficients. Some assembly districts had nearly twice as large associations compared to the overall mean within short distances. Some assembly districts also had associations that decrease faster towards zero, making it plausible that each assembly district may have its own effective buffer size.

The assembly district-specific average buffer effects up to 1/2 mile, $\bar{\beta}_j(0; 1/2)$, are given in Figure 4.2c. We found that some districts have a significant positive 1/2 mile buffer effect. Since those significant positive 1/2 mile buffer effects are related to having significant DL effects around 1/2 miles, we also examined if credible intervals of estimated DL effects overlap with 0 along distances up to 7 miles for each assembly district (Figure 4.3). We found that some assembly districts had significant DL effects

Figure 4.2: **Estimated (a) overall and (b) assembly district specific DL effects, and (c) estimated 1/2 mile buffer effects of assembly districts from model 3 adjusted for individual characteristics. Data sources: 2001-2010 Fitnessgram and NETS databases.**

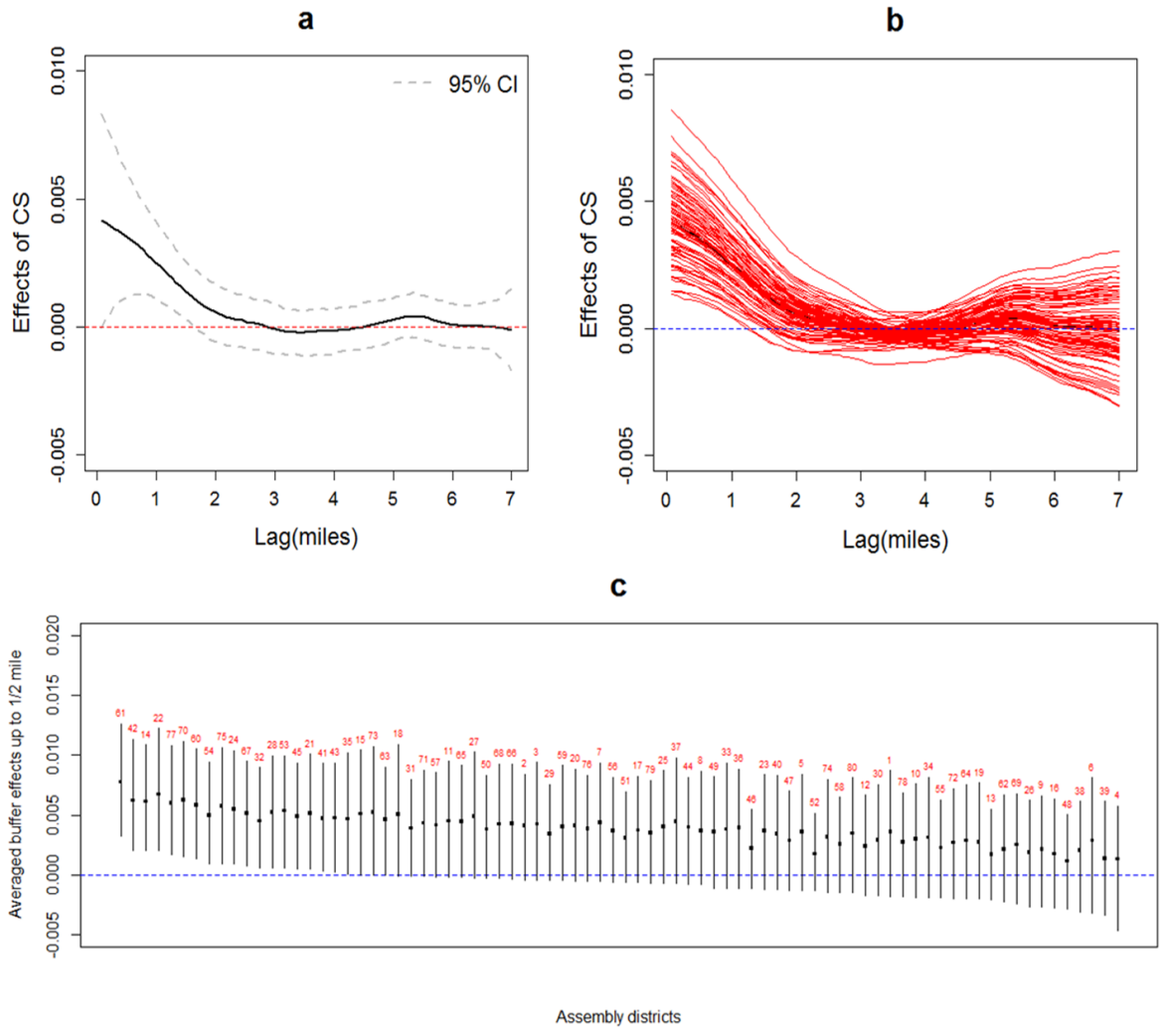
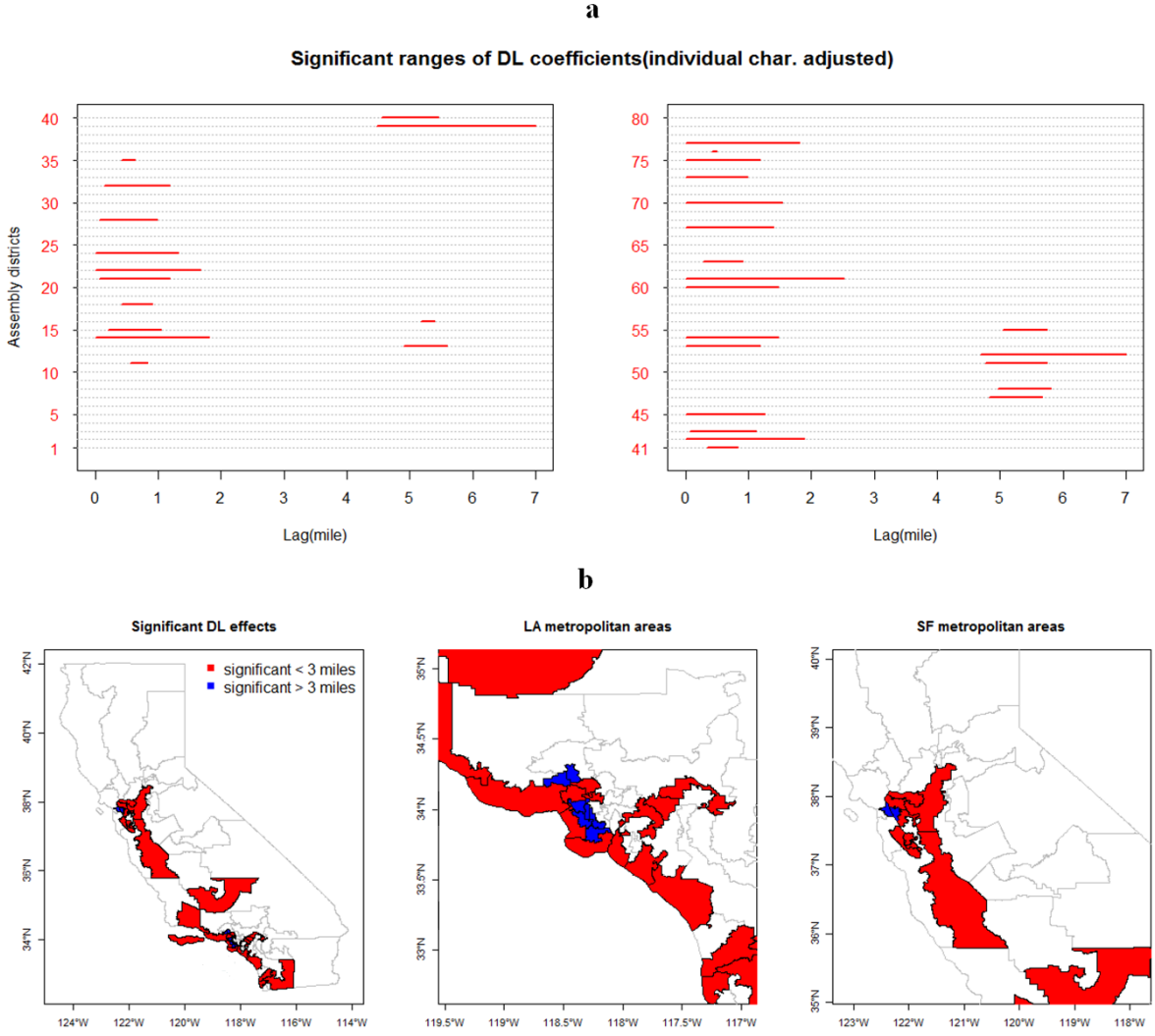


Figure 4.3: **(a) Distances at which credible intervals of DL effects up do not overlap with 0 across California's 80 assembly districts. (b) Mapped assembly districts with significant DL effects before (red) and after (blue) 3 miles.**



before or after 3 miles (Figure 4.3a). Assembly districts with significant DL effects tend to be in more highly urbanized areas (Figure 4.3b); primarily regions surrounding San Francisco and Los Angeles. Within these regions, districts with significant DL effects within 3 miles are mostly located in suburban areas, whereas inner city areas have significant associations after 3 miles. Degree of urbanization (and its correlates such as income) may be major drivers of the magnitude of the associations as well as the distances from schools that are most influential.

In additional analyses we included adjustment for other school and school-neighborhood

characteristics: total student enrollment in a school, percentage of children in a school that participated in the California school free or reduced meal program, and percentage of adults with a bachelor's degree or higher residing within schools' census tracts. As may be expected due to potential confounding and/or mediation (*Chaix et al.*, 2010; *Diez Roux*, 2004), these adjustments attenuated the overall and district-specific DL coefficients. The number of assembly districts with significant DL effects before or after 3 miles decreased, but there were still some assembly districts with DL coefficients that remained significant (See Appendix E Figure E.4-E.5).

4.6 Discussion

We systematically examined variability in the associations between the presence of convenience stores near schools, a measure of the built food environment, and children's weight using HDLMs. We found differences in the associations at the California assembly district level. First, some assembly districts had about 2 fold higher associations than the overall mean. Second, there was some indication that the distances at which the associations are significant varies across districts. These findings suggest that in certain assembly districts children's body weight may be more vulnerable to aspects of the built food environment.

This is the first study to systematically and comprehensively evaluate differences in health-built environment associations across a large geographical area, namely California, the most populous and diverse state in the United States. We capitalized on the availability of the FitnessGram data for the years 2001-2010, a surveillance database of over 3.6 million 7^{th} grade public school children which contains measured BMIz, and the comprehensive NETS database from which features of the built environment were derived.

HDLMs with varying degrees of complexity were constructed based upon exploratory analyses; they were compared using DIC as a way to describe the types of

variations in the district-specific DL effects. We further examined the possible spatial correlation of the district effects which might have been caused by spatially structured unobserved covariates. Neither the average BMIz nor the built-environment associations were spatially patterned; this suggests that more localized differences in environments or policies may explain the observed differences across assembly districts. When it is necessary to have spatial HDLMs due to lack of observed spatial information, spatially structured random intercepts and DL coefficients could be similarly modeled as in *Macnab and Gustafson* (2007).

The proposed multilevel modeling approach can serve more broadly to help advance the built environment-health literature to incorporate complex data structures. More importantly however, this approach can help address scientific questions regarding the role of the built environment that have long been speculated yet not systematically investigated. Specifically, these methods provide a way to investigate the underlying overall shape of built environment-health associations along distances, as well as area-level heterogeneity of those associations.

CHAPTER V

Conclusion and Future Work

Built environment researchers often encounter complex data structures in their research. For instance, data may be collected over time and space, observations may be nested in groups or regions, and the underlying spatial mechanism of how factors in the built environment affect health is usually unknown. These complex data give rise to several data analysis challenges, since assessing associations of built environment factors on health with valid inferences requires construction of statistical models that incorporate various sources of variations that come from many factors: for example, individual characteristics (including age, gender, race/ethnicity, income, and level of education) and neighborhood characteristics (such as accessibility to health services, availability of food stores, walkability of neighborhoods, and food quality). Without observing all the information relevant to the association between factors of the built environment and health, statistical modeling to account for such uncertainty is important to make valid inferences for measured built environment effects. To improve traditional approaches commonly implemented in the built environment literature, this dissertation has focused on developing novel statistical methodologies for examining associations between factors of the built environment and health outcomes.

In Chapter 2 as an extension of multiple informant models (*Pepe et al.*, 1999; *Horton et al.*, 1999), a hierarchical multiple informant model (HMIM) was developed

to examine or test whether marginal associations of types of the built environment factors have a different impact on a health outcome while accounting for a hierarchical data structure (i.e., children are nested in schools). Since HMIMs are built upon generalized estimating equation (GEE) methods, all the properties of HMIMs are induced by GEE, such as consistency of parameter estimates and robustness of standard errors of regression parameter estimates. Additionally, any available statistical software for GEE can easily be used to fit HMIMs for practical applicability.

As noted in the discussion of Chapter 2 for possible extensions of the work, in Chapter 3 we examined how associations of measured built environment factors on health change over distance from outcome locations through distributed lag models (DLMs). We found that DLMs enable us not only to examine the shape of the built environment associations across distance from study locations, but also to estimate buffer effects (e.g., a buffer effect up to r mile from the outcome locations) of built environment factors; furthermore, these estimated buffer effects were more accurate than those from traditional linear models under various conditions of spatial clustering of factors in the built environment.

However, this clustering of the built environment factors may suggest that propensities of being exposed to the built environment factors are not equal for all the observations in the study. If a question of interest is estimating buffer effects up to a certain distance from locations of interest, a propensity score method may be considered to account for the unbalanced built environment exposures. Specifically, if we have information that explains clustering of the built environment factors, we may use that additional information to allocate different propensities to observations and estimate buffer effects from traditional linear models. This is a promising direction for future work.

In Chapter 4, DLMs were further extended to describe how associations between features of the built environment and health vary by areas due to unobserved areal

characteristics or inherently different associations. Through a hierarchical distributed lag models (HDLM), we examined the possibility of different areas having different relevant buffer sizes within which measured built environment factors may have health effects, and located areas where subjects were more vulnerable to exposures of the built environment.

However, the HDLM implemented in Chapter 4 still assumes that all the units within the same area have the same effective buffer size and patterns; in fact, there is a possibility that each unit of observation may have different buffer sizes due to unobserved covariates. A possible approach that examines different buffer sizes of measured built environment factors on health across units may be through a modification of spatially varying coefficient models (*Gelfand et al.*, 2003).

In the case that covariates are random or partially observed with a spatial pattern, it may be hard to use spatially varying coefficient models because both coefficients and covariates are unobserved, as noted by *Heaton and Gelfand* (2011). That is, untangling spatial parameters between coefficients and covariates may be difficult. However, in the case of completely observed covariates that are spatially correlated, it may be possible to consistently estimate spatially varying coefficients if we have information that explains the spatial pattern of observed covariates. For instance, spatial parameters in spatially varying coefficients may be consistently estimable if we adjust the propensities of being exposed to the built environment factors across units, as suggested for further work building on Chapter 4.

This dissertation has provided novel directions and methods with which to study how built environment factors affect health. While several challenges and future work remain, through the proposed methodologies we have shown how to begin to answer questions long posed in the built environment health effects literature.

APPENDICES

APPENDIX A

R code for empirical covariance matrix

```
1
2  ## data ##
3  ## data need sorted by id, beforehand
4  names(data) = c("id", "y", "x1", "x2")
5  ## use gee library for gee function
6  library(gee)
7  gee.model1 = gee(y ~ x1, id = id, data=data, corstr = "exchangeable")
8  gee.model2 = gee(y ~ x2, id = id, data=data, corstr = "exchangeable")
9
10 ## Total number of observations
11 N = gee.model1$nobs
12
13 ## Number of parameters from each marginal GEE model
14 p = length(gee.model1$coefficients)
15
16 ## number of subjects in cluster
17 n_j = table(gee.model1$id)
18
19 ## the number of multiple informants
20 n_x = 2
21
22 ## correlation from each fitted GEE model ##
23 rho_1 = gee.model1$working.correlation[1,2]
24 rho_2 = gee.model2$working.correlation[1,2]
25
26 ## dispersion from each fitted GEE model ##
```

```
phi_1 = gee.model1$scale
phi_2 = gee.model2$scale

## bread for each fitted GEE model ##
bread_1 = gee.model1$naive.variance
bread_2 = gee.model2$naive.variance

## create design matrix of each fitted GEE model ##
X1 = model.matrix( ~ x1, data=data)
X2 = model.matrix( ~ x2, data=data)

## initializing for big filling ##
big.filling = matrix(0, nrow = n_x*p, ncol = n_x*p)

## initializing for indexing clusters ##
s = 0
## calculate for big filling ##
for(i in 1:nrow(n_j)){
  V_1 = matrix(rho_1*phi_1, nrow = n_j[i], ncol = n_j[i])
  V_2 = matrix(rho_2*phi_2, nrow = n_j[i], ncol = n_j[i])

  diag(V_1) = phi_1
  diag(V_2) = phi_2

  ## define index ##
  index = seq(1+s, length = n_j[i])

  tmp_1 = t(X1[index,]) %*% solve(V_1) %*% gee.model1$residuals[index]
  tmp_2 = t(X2[index,]) %*% solve(V_2) %*% gee.model2$residuals[index]

  big.tmp = rbind(tmp_1, tmp_2)

  big.filling = big.filling + big.tmp %*% t(big.tmp)
  s = s + n_j[i]
}

## use magic library for a block diagonal matrix using adiag function ##
library(magic)

## make big bread ##
big.bread = adiag(bread_1, bread_2)
```

```
68
69 ## make a sandwich or empirical variance estimator ##
70 emp.variance = big.bread %*% big.filling %*% big.bread
71 emp.variance
```

APPENDIX B

SWEEP operator

The Sweep operator is a linear operator that works as follows to produce marginal or conditional effects of a given predictor on the outcome. Let the marginal mean vector be $\boldsymbol{\mu}^T = [\ \mu_{X_1} \quad \mu_{X_2} \quad \mu_{X_3} \quad \mu_Y\ ]$, and $\boldsymbol{\Sigma}$ is covariance matrix of (X_1, X_2, X_3, Y). Then, sweeping $\begin{bmatrix} -1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu}^T & \boldsymbol{\Sigma} \end{bmatrix}$ on the row and column of X_k yields the marginal effects from the regression of Y on X_k. For instance, sweeping $\begin{bmatrix} -1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu}^T & \boldsymbol{\Sigma} \end{bmatrix}$ on the row and column of X_3 provides the equality below:

$$SWP[X_3] \begin{bmatrix} -1 & \mu_{X_1} & \mu_{X_2} & \mu_{X_3} & \mu_Y \\ \cdot & \sigma^2_{X_1} & \sigma_{X_1X_2} & \sigma_{X_1X_3} & \sigma_{YX_1} \\ \cdot & \cdot & \sigma^2_{X_2} & \sigma_{X_2X_3} & \sigma_{YX_2} \\ \cdot & \cdot & \cdot & \sigma^2_{X_3} & \sigma_{YX_3} \\ \cdot & \cdot & \cdot & \cdot & \sigma^2_Y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} = \beta_{03} \\ \cdot & a_{22} & a_{23} & a_{24} & a_{25} \\ \cdot & \cdot & a_{33} & a_{34} & a_{35} \\ \cdot & \cdot & \cdot & a_{44} & a_{45} = \beta_{13} \\ \cdot & \cdot & \cdot & \cdot & a_{55} = Var[Y|X_3] \end{bmatrix},$$

where a_{rc} is the row r and column c element of the swept matrix of $\begin{bmatrix} -1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu}^T & \boldsymbol{\Sigma} \end{bmatrix}$ on the row and column of X_3. Note that since the above matrices are symmetric, we only

recorded upper triangular elements. From the result, the swept matrix on X_3 yields the marginal effects β_{03}, β_{13} from the regression of Y on X_3. Similarly, sweeping the matrix $\begin{bmatrix} -1 & \boldsymbol{\mu}^T \\ \boldsymbol{\mu}^T & \boldsymbol{\Sigma} \end{bmatrix}$ on X_1, X_2, and X_3, results in

$$SWP[X_1\, X_2\, X_3] \begin{bmatrix} -1 & \mu_{X_1} & \mu_{X_2} & \mu_{X_3} & \mu_Y \\ \cdot & \sigma^2_{X_1} & \sigma_{X_1X_2} & \sigma_{X_1X_3} & \sigma_{YX_1} \\ \cdot & \cdot & \sigma^2_{X_2} & \sigma_{X_2X_3} & \sigma_{YX_2} \\ \cdot & \cdot & \cdot & \sigma^2_{X_3} & \sigma_{YX_3} \\ \cdot & \cdot & \cdot & \cdot & \sigma^2_Y \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} = \gamma_0 \\ \cdot & b_{22} & b_{23} & b_{24} & b_{25} = \gamma_1 \\ \cdot & \cdot & b_{33} & b_{34} & b_{35} = \gamma_2 \\ \cdot & \cdot & \cdot & b_{44} & b_{45} = \gamma_3 \\ \cdot & \cdot & \cdot & \cdot & b_{55} = \sigma^2_\delta \end{bmatrix}.$$

Sweeping operators were used for the marginal moments at the cluster level to derive cluster-level mean and variances.

APPENDIX C

True marginal covariance structure in the simulation study

Let Y_{ij} be an outcome of the i^{th} child, $i = 1, 2, \ldots, n_j$, in the j^{th} school, $j = 1, 2, \ldots, J$. With three covariates, X_{j1}, X_{j2}, and X_{j3}, the outcome Y_{ij} is generated from the following hierarchical model

$$Y_{ij} = \gamma_0 + \gamma_1 X_{j1} + \gamma_2 X_{j2} + \gamma_3 X_{j3} + \delta_j + \epsilon_{ij},$$

where $\delta_j \sim N(0, \sigma_\delta^2)$ and ϵ_{ij} $N(0, \sigma_\epsilon^2)$.

Suppose that we fit a HMIM to compare marginal effects among three covariates X_{j1}, X_{j2}, and X_{j3} such that $Var[X_{jk}|X_{jk'}] = \sigma^2_{k|k'}$ for $k, k' = 1, 2, 3, k \neq k'$, and $Cov(X_{j2}|X_{j1}, X_{j3}|X_{j1}) = \sigma_{23|1}, Cov(X_{j1}|X_{j3}, X_{j2}|X_{j3}) = \sigma_{12|3}$, and $Cov(X_{j1}|X_{j2}, X_{j3}|X_{j2}) = \sigma_{13|2}$.

Then, the marginal variance and correlation can be written as

$$\begin{aligned}
Var[Y_{ij}|X_{j1}] &= \gamma_2^2\sigma^2_{2|1} + \gamma_3^2\sigma^2_{3|1} + 2\gamma_2\gamma_3\sigma_{23|1} + \sigma_\delta^2 + \sigma_\epsilon^2 \\
Var[Y_{ij}|X_{j2}] &= \gamma_1^2\sigma^2_{1|2} + \gamma_3^2\sigma^2_{3|2} + 2\gamma_1\gamma_3\sigma_{13|2} + \sigma_\delta^2 + \sigma_\epsilon^2 \\
Var[Y_{ij}|X_{j3}] &= \gamma_1^2\sigma^2_{1|3} + \gamma_2^2\sigma^2_{2|3} + 2\gamma_1\gamma_2\sigma_{12|3} + \sigma_\delta^2 + \sigma_\epsilon^2
\end{aligned}$$

and

$$\begin{cases} Corr(Y_{ij}|X_{j1}, Y_{i'j}|X_{j1}) = (\gamma_2^2\sigma_{2|1}^2 + \gamma_3^2\sigma_{3|1}^2 + 2\gamma_2\gamma_3\sigma_{23|1} + \sigma_\delta^2)/Var[Y_{ij}|X_{j1}] \\ Corr(Y_{ij}|X_{j2}, Y_{i'j}|X_{j2}) = (\gamma_1^2\sigma_{1|2}^2 + \gamma_3^2\sigma_{3|2}^2 + 2\gamma_1\gamma_3\sigma_{13|2} + \sigma_\delta^2)/Var[Y_{ij}|X_{j2}] \\ Corr(Y_{ij}|X_{j3}, Y_{i'j}|X_{j3}) = (\gamma_1^2\sigma_{1|3}^2 + \gamma_2^2\sigma_{2|3}^2 + 2\gamma_1\gamma_2\sigma_{12|3} + \sigma_\delta^2)/Var[Y_{ij}|X_{j3}] \end{cases}$$

Then, the true covariance structure setting of the HMIM in the simulation section is derived as

$$\tilde{\mathbf{V}}_j = \begin{bmatrix} \mathbf{V}_{j1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{j2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_{j3} \end{bmatrix},$$

where for $k = 1, 2, 3$, $\mathbf{V}_{jk} = \phi\mathbf{R}_{jk}$ with $\mathbf{R}_{jk} = \begin{bmatrix} 1 & \rho_k & \cdots & \rho_k \\ \rho_k & 1 & \cdots & \rho_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_k & \cdots & 1 \end{bmatrix}$, and $\phi_k = Var[Y_{ij}|X_{jk}]$ and $\phi_k = Corr(Y_{ij}|X_{jk}, Y_{i'j}|X_{jk})$.

APPENDIX D

Parameter estimation in DLMs

In a matrix form, Eq. (3.2) can be written as $\boldsymbol{\beta} = \boldsymbol{C_0\alpha} + \boldsymbol{C_1\tilde{\alpha}}$, where $\boldsymbol{C}_0 = \begin{bmatrix} 1 & r_1 \\ \vdots & \vdots \\ 1 & r_L \end{bmatrix}, \boldsymbol{C}_1 = [|r_l - r_k|^3]_{1\leq l,k\leq L}, \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$, and $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_L)^T$. The coefficients $\tilde{\boldsymbol{\alpha}}$ are penalized so the squared second derivative of the estimated DL coefficient function is penalized. The objective is to minimize $||\mathbf{Y} - \mathbf{1}_n\beta_0 - \mathbf{X}\boldsymbol{\beta}||^2 = ||\mathbf{Y} - \mathbf{1}_n\boldsymbol{\beta}_0 - \mathbf{X}(\boldsymbol{C}_0\boldsymbol{\alpha} + \boldsymbol{C}_1\tilde{\boldsymbol{\alpha}})||^2$ subject to the constrains $\tilde{\boldsymbol{\alpha}}^T\boldsymbol{C}_1\tilde{\boldsymbol{\alpha}} \leq const$, and $\boldsymbol{C}_0^T\tilde{\boldsymbol{\alpha}} = 0$. The latter constraint implies that there are really L free parameters $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\alpha}}$ rather than $L+2$ implied from the columns of $\boldsymbol{C}_0$ and $\boldsymbol{C}_1$ (*Green and Silverman*, 1993; *Ruppert et al.*, 2003). As is well known, the optimization problem can be re-written as a mixed model by redefining $\tilde{\boldsymbol{\alpha}} = \mathbf{M}_1\boldsymbol{a}_1$, where $\mathbf{M}_1$ is an $L \times (L-2)$ orthogonal matrix to $\boldsymbol{C}_0$, where $\mathbf{M}_1$ can be determined using the QR decomposition $[\boldsymbol{C}_0\ \boldsymbol{C}_1] = \mathbf{Q}_c\mathbf{R}_c$ and setting $\mathbf{M}_1$ as the 3^{rd} to last columns of $\mathbf{Q}_c$(*Green and Silverman*, 1993). Further, finding $\mathbf{M}_2^{1/2}$ that satisfies $\mathbf{M}_2 = \mathbf{M}_2^{1/2}\mathbf{M}_2^{1/2} = \mathbf{M}_1^T\boldsymbol{C}_1\mathbf{M}_1$, and defining $\boldsymbol{b}_1$ through the transformation $\boldsymbol{a}_1$ to $\mathbf{M}_2^{-1/2}\boldsymbol{b}_1$, and re-structuring the data $\mathbf{X}^* = [\mathbf{1}_n\ \mathbf{X}\boldsymbol{C}_0]$ and $\mathbf{Z}^* = \mathbf{X}\boldsymbol{C}_1\mathbf{M}_1\mathbf{M}_2^{-1/2}$, the mixed model becomes $\mathbf{Y} = \mathbf{X}^*[\beta_0\ \boldsymbol{\alpha}^T] + \mathbf{Z}^*\boldsymbol{b}_1 + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \tau^2\boldsymbol{I})$ and $\boldsymbol{b}_1 \sim N_{L-2}(\mathbf{0}, \sigma_b^2\boldsymbol{I})$. The smoothing parameter is $\lambda = \tau^2/\sigma_b^2$.

The mixed model can be fitted with packaged software for mixed models in the frequentist framework. Once we have the estimates from the fitted regression, the estimates of the DL coefficients can be obtained as $\boldsymbol{\beta} = \boldsymbol{\Omega}\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{b}_1 \end{bmatrix}$ and $Cov(\boldsymbol{\beta}) = \boldsymbol{\Omega} Cov(\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{b}_1 \end{bmatrix})\boldsymbol{\Omega}^T$ where $\boldsymbol{\Omega} = [\boldsymbol{C}_0\ \boldsymbol{C}_1\mathbf{M}_1\mathbf{M}_2^{-1/2}]$.

Alternatively, the model can be estimated in the Bayesian framework. With prior distributions of $\beta_0 \propto 1, \boldsymbol{\alpha} \propto \mathbf{1}, \boldsymbol{b}_1 \sim N(\mathbf{0}, \sigma_b^2 \boldsymbol{I}_{L-2}), \sigma_b^2 \sim IG(a_\sigma, b_\sigma), \tau^2 \sim IG(a_\tau, b_\tau)$, the full conditionals are all available in closed forms. Let $\mathbf{D}^* = [\mathbf{X}^*\ \mathbf{Z}^*] = [\mathbf{1}_n\ \mathbf{X}\boldsymbol{C}_0\ \mathbf{X}\boldsymbol{C}_1\mathbf{M}_1\mathbf{M}_2^{-1/2}]$, then the full conditional for $\beta_0, \boldsymbol{\alpha}, \boldsymbol{b}_1$ is $p(\beta_0, \boldsymbol{\alpha}, \boldsymbol{b}_1|\cdot) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (\mathbf{D}^{*T}\mathbf{D}^*/\tau^2 + \sigma_b^{-2}\mathbf{G})^{-1}, \mathbf{G} = diag\{\mathbf{0}_3, \mathbf{1}_{L-2}\}$ and $\boldsymbol{\mu} = \boldsymbol{\Sigma}\mathbf{D}^{*T}\mathbf{Y}/\tau^2$. The full conditional distribution for σ_b^2 is $p(\sigma_b^2|\cdot) = IG(a_\sigma + (L-2)/2, b_\sigma + \boldsymbol{b}_1^T\mathbf{G}\boldsymbol{b}_1/2)$, while the full conditional distribution of τ^2 is $p(\tau^2|\cdot) = IG(a_\tau + n/2, b_\tau + (\boldsymbol{r}^T\boldsymbol{r})/2)$, where $\boldsymbol{r} = \mathbf{Y} - \mathbf{D}^*(\beta_0, \boldsymbol{\alpha}, \boldsymbol{b}_1)^T$. Inference for DL coefficients $\boldsymbol{\beta}$ is obtained by transforming posterior samples of $\boldsymbol{\alpha}, \boldsymbol{b}_1$ by $\boldsymbol{\Omega}\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{b}_1 \end{bmatrix}$ with $\boldsymbol{\Omega}$ as described above. Inference for average lag effects, $\bar{\beta}(0; r_k)$, can be easily determined from posterior samples.

Figure D.1: **The built environment setting; locations of food stores are sampled from an inhomogeneous Poisson point process (a) without clustering, (b) with a small amount of clustering, (c) with a large amount of clustering.**

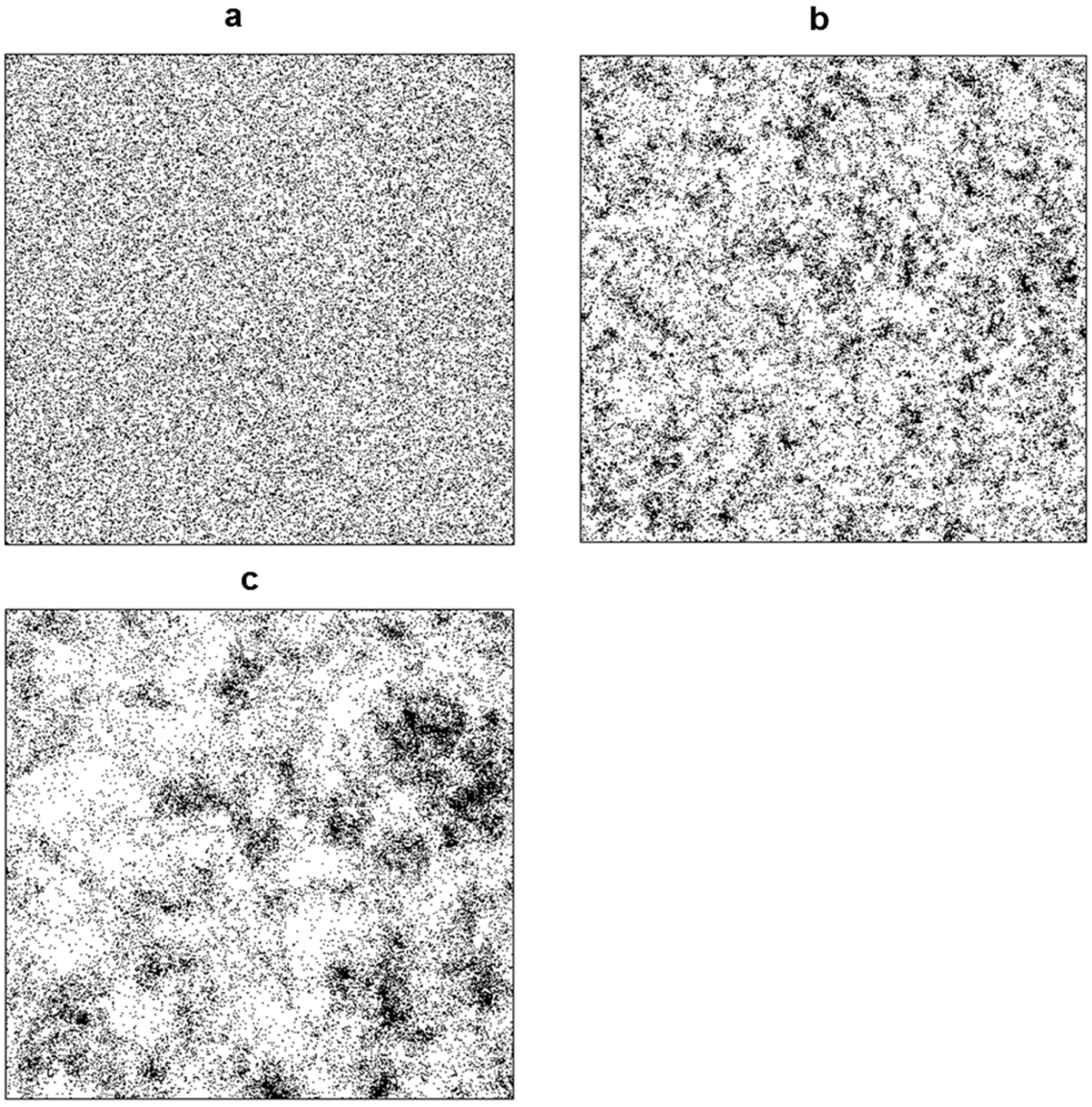


Figure D.2:

Bias, variance, MSE, and coverage rate at each $r_l, l = 1, 2, \ldots, 100$ for the cases when $\beta(r)$ is: (a) a step function under the built environment without clustering. (b) the step function under the built environment with a large amount of clustering. (c) $\beta(r)$ is the normal *pdf* under the built environment without clustering. (d) $\beta(r)$ is the normal *pdf* under the built environment with a large amount of clustering. Reported results are from a simulation case with $n = 1,000$ and $R^2 = 0.2$.

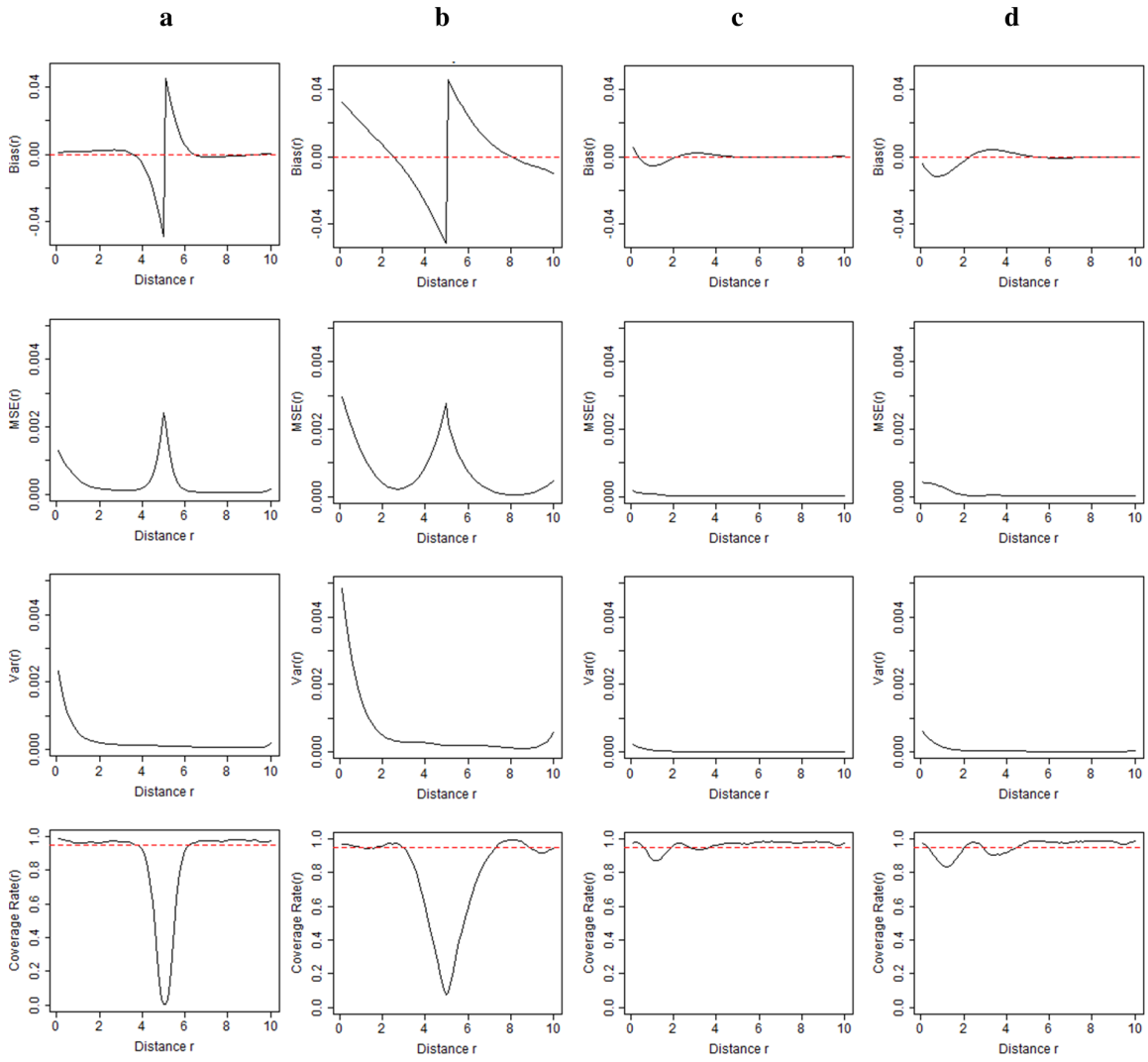


Table D.1: **Simulation results for the averaged buffer effects up to distance r_k = 2.5, 5, and 7.5 from the traditional model and the fitted DLM. Reported results are from a simulation case with n=1,000 and R^2=0.2.**

			r_k = 2.5		r_k = 5		r_k = 7.5	
$\beta(r)$	Fitted model	Spatial range in the built environment	Est. beta	Coverage rate	Est. beta	Coverage rate	Est. beta	Coverage rate
Step		True $\bar{\beta}(0; r_k)$	0.100	-	0.100	-	0.044	-
	Traditional linear models	Independence	0.102	0.942	0.100	0.941	0.044	0.956
		5	0.242	0.000	0.100	0.947	0.051	0.600
		20	0.320	0.000	0.100	0.944	0.047	0.847
	DLM	Independence	0.102	0.954	0.091	0.786	0.045	0.956
		5	0.111	0.934	0.085	0.573	0.047	0.908
		20	0.110	0.945	0.083	0.663	0.047	0.921
Curve		True $\bar{\beta}(0; r_k)$	0.058	-	0.021	-	0.010	-
	Traditional linear models	Independence	0.058	0.946	0.021	0.948	0.009	0.946
		5	0.074	0.134	0.024	0.685	0.011	0.526
		20	0.078	0.036	0.022	0.863	0.010	0.732
	DLM	Independence	0.056	0.919	0.021	0.947	0.009	0.953
		5	0.055	0.900	0.022	0.949	0.009	0.954
		20	0.054	0.889	0.022	0.959	0.010	0.959

Table D.2: **Integrated MSE from fitted traditional linear models with distance lag r_k = 2.5, 5, and 7.5 and from fitted DLMs with a maximum distance r_L = 10. Reported results are from a simulation case with $n = 1,000$ and R^2=0.2.**

		Traditional linear model (r_k=2.5)	Traditional linear model (r_k=5)	Traditional linear model (r_k=7.5)	DLM
$\beta(r)$	Spatial range in the built environment	IMSE*	IMSE*	IMSE*	IMSE*
Step	Independence	2.568	0.026	2.077	0.321
	5	7.646	0.022	1.872	0.613
	20	14.703	0.016	1.975	0.755
Curve	Independence	0.213	0.796	1.117	0.021
	5	0.175	0.745	1.061	0.040
	20	0.184	0.769	1.084	0.064

* IMSE is multiplied by 100.

APPENDIX E

Parameter estimation in HDLMs

In a matrix form, Eq. (4.2) can be written as $\boldsymbol{\beta} = \boldsymbol{C_0\alpha} + \boldsymbol{C_1\tilde{\alpha}}$, where $\boldsymbol{C}_0 = \begin{bmatrix} 1 & r_1 \\ \vdots & \vdots \\ 1 & r_L \end{bmatrix}, \boldsymbol{C}_1 = [|r_l - r_k|^3]_{1\leq l,k\leq L}, \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}$, and $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_L)^T$. The coefficients $\tilde{\boldsymbol{\alpha}}$ are penalized so the squared second derivative of the estimated DL coefficient function is penalized. The objective is to minimize $||\mathbf{Y} - \mathbf{1}_n\beta_0 - \mathbf{X}\boldsymbol{\beta}||^2 = ||\mathbf{Y} - \mathbf{1}_n\boldsymbol{\beta}_0 - \mathbf{X}(\boldsymbol{C}_0\boldsymbol{\alpha} + \boldsymbol{C}_1\tilde{\boldsymbol{\alpha}})||^2$ subject to the constrains $\tilde{\boldsymbol{\alpha}}^T\boldsymbol{C}_1\tilde{\boldsymbol{\alpha}} \leq const$, and $\boldsymbol{C}_0^T\tilde{\boldsymbol{\alpha}} = 0$. The latter constraint implies that there are really L free parameters $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\alpha}}$ rather than $L+2$ implied from the columns of $\boldsymbol{C}_0$ and $\boldsymbol{C}_1$ (*Green and Silverman*, 1993; *Ruppert et al.*, 2003). As is well known, the optimization problem can be re-written as a mixed model by redefining $\tilde{\boldsymbol{\alpha}} = \mathbf{M}_1\boldsymbol{a}_1$, where $\mathbf{M}_1$ is an $L \times (L-2)$ orthogonal matrix to $\boldsymbol{C}_0$, where $\mathbf{M}_1$ can be determined using the QR decomposition $[\boldsymbol{C}_0\ \boldsymbol{C}_1] = \mathbf{Q}_c\mathbf{R}_c$ and setting $\mathbf{M}_1$ as the 3^{rd} to last columns of $\mathbf{Q}_c$(*Green and Silverman*, 1993). Further, finding $\mathbf{M}_2^{1/2}$ that satisfies $\mathbf{M}_2 = \mathbf{M}_2^{1/2}\mathbf{M}_2^{1/2} = \mathbf{M}_1^T\boldsymbol{C}_1\mathbf{M}_1$, and defining $\boldsymbol{b}_1$ through the transformation $\boldsymbol{a}_1$ to $\mathbf{M}_2^{-1/2}\boldsymbol{b}_1$, and re-structuring the data $\mathbf{X}^* = \mathbf{X}\boldsymbol{C}_0$ and $\mathbf{Z}^* = \mathbf{X}\boldsymbol{C}_1\mathbf{M}_1\mathbf{M}_2^{-12}$.

Then, the mixed model only with a random intercept and random DL effects of

assembly districts becomes

$$Y_{ijt} = \beta_0 + \eta_j + \mathbf{X}^*_{ijt}(\boldsymbol{\alpha} + \boldsymbol{\alpha}_j) + \mathbf{Z}^*_{ijt}(\boldsymbol{b}_1 + \boldsymbol{b}_{1j}) + \epsilon_{ijt}, \tag{E.1}$$

where $\eta_j \sim N(0, \tau^2_\eta), \boldsymbol{\alpha}_j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\alpha), \boldsymbol{b}_1 \sim N_{L-2}(\mathbf{0}, \sigma^2_{b_1}\mathbf{I}), \boldsymbol{b}_{1j} \sim N_{L-2}(\mathbf{0}, \sigma^2_{b_2}\mathbf{I})$, and $\epsilon_{ijt} \sim N(0, \tau^2)$. With integrating over $(\eta_j, \boldsymbol{\alpha}_j, \boldsymbol{b}_{1j})$, the marginal model of Y_{ijt} becomes

$$Y_{ijt} = \beta_0 + \mathbf{X}^*_{ijt}\boldsymbol{\alpha} + \mathbf{Z}^*_{ijt}\boldsymbol{b}_1 + \epsilon^{'}_{ijt}, \tag{E.2}$$

where $\epsilon^{'}_{ijt} \sim N(0, \tau^2_\eta + \mathbf{X}^*_{ijt}\boldsymbol{\Sigma}_\alpha\mathbf{X}^{*T}_{ijt} + \sigma^2_{b_2}\mathbf{Z}^*_{ijt}\mathbf{Z}^{*T}_{ijt} + \tau^2)$.

The sampling approach we use is based on *Agarwal and Gelfand* (2005) shrinkage slice sampling methods which is an extension of *Neal* (2003). Slice sampling refers to using auxiliary variables to draw posterior samples where samples are not easily drawn, and shrinkage bounds the sampling domain based on the rejected proposal samples. The sampling steps are

(a) Partition the parameters into of $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\alpha}, \boldsymbol{b}_1)$ and $\boldsymbol{\Omega} = (\tau^2_\eta, \boldsymbol{\Sigma}_\alpha, \sigma^2_{b_2}, \tau^2)$ so that $f(\boldsymbol{\theta}|\boldsymbol{\Omega}, Y)$ is easy to sample from closed form of the full conditional.

(b) Start with initial values of parameters $\boldsymbol{\theta}^0$ and $\boldsymbol{\Omega}^0$.

(c) Draw θ from $f(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{Y})$, given below.

Implement the shrinkage slice sampling steps

(d) Draw $\nu = -l(\boldsymbol{\theta}, \boldsymbol{\Omega}|Y) + z$, where $z \sim exp(1)$.

(e) Draw $\boldsymbol{\Omega}$ from $\pi(\boldsymbol{\Omega}|\boldsymbol{\theta}, \nu)I(-l(\boldsymbol{\theta}, \boldsymbol{\Omega}|Y) < \nu < \infty)$, given below.

(f) Iterate (c) through (e) until we get the appropriate number of MCMC samples.

The density $f(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{Y})$ in (c) is derived as follows. Let $\boldsymbol{\Psi} = \tau^2_\eta\mathbf{U}_B\mathbf{U}^T_B + \mathbf{X}^*_B\boldsymbol{\Sigma}_B\mathbf{X}^{*T}_B + \sigma^2_{b_2}\mathbf{Z}^*_B\mathbf{Z}^{*T}_B + \tau^2\mathbf{I}$ be the covariance matrix of $\mathbf{Y}$ marginalized in (E.2), where $\mathbf{U}_B$ is a block diagonal matrix of $\mathbf{1}_{n_j}$ for assembly district $j = 1, 2, \ldots, 80$, $\mathbf{X}^*_B$ and $\mathbf{Z}^*_B$ are a block diagonal matrix of district's own $\mathbf{X}^*_j$ and $\mathbf{Z}^*_j$, $\boldsymbol{\Sigma}_B = diag\{\boldsymbol{\Sigma}_\alpha, \ldots, \boldsymbol{\Sigma}_\alpha\}$, and let $\mathbf{D}^* = [\mathbf{1}_n\ \mathbf{X}^*\ \mathbf{Z}^*]$. Then the posterior distribution of $\boldsymbol{\theta}$ is $f(\boldsymbol{\theta}|\boldsymbol{\Omega}, \mathbf{Y}) = N(\mathbf{S}^{-1}\mathbf{m}, \mathbf{S}^{-1})$,

where $\mathbf{S} = (\mathbf{D}^{*T}\mathbf{\Psi}^{-1}\mathbf{D}^{*} + diag\{0,0,0,\sigma_b^{-2}\mathbf{I}_{L-2}\})$ and $\mathbf{D}^{*T}\mathbf{\Psi}^{-1}\mathbf{Y}$.

In step (d), each parameter of $\mathbf{\Omega} = (\tau_\eta^2, \mathbf{\Sigma}_\alpha, \sigma_{b_2}^2, \tau^2)$ can be updated one at a time or simultaneously, we updated one at a time. Without loss of generality, we draw a sample for τ_η^2 first. Given the sampled $\boldsymbol{\theta}^{(t)}$, we evaluate the log-likelihood $l(\boldsymbol{\theta}^{(t)}, \mathbf{\Omega}^{(t-1)}|\mathbf{Y})$ and draw $\nu_1 = -l(\boldsymbol{\theta}^{(t)}, \mathbf{\Omega}^{(t-1)}|\mathbf{Y}) + z_1$, where $z_1 \sim exp(1)$. Then, sample $\tau_\eta^{2(t)}$ from the prior inverse gamma distribution $IG(a_\eta, b_\eta)$ with bounds $L_\eta < \tau_\eta^{2(t)} < U_\eta$, where L_η and U_η are lower and upper bounds of τ_η^2, respectively, and evaluate $l(\boldsymbol{\theta}^{(t)}, \tau_\eta^{2(t)}, \mathbf{\Omega}_{-\tau_\eta^2}^{(t-1)}|\mathbf{Y})$. The posterior sample $\tau_\eta^{2(t)}$ is accepted when $-l(\boldsymbol{\theta}^{(t)}, \tau_\eta^{2(t)}, \mathbf{\Omega}_{-\tau^2}^{(t-1)}|\mathbf{Y}) < \nu_1$. If $\tau_\eta^{2(t)}$ is rejected, the bounds of the prior distribution are shrunk such that $L_\eta = \tau_\eta^{2(t)}$ if $\tau_\eta^{2(t)} < \tau_\eta^{2(t-1)}$ or $U_\eta = \tau_\eta^{2(t)}$ if $\tau_\eta^{2(t)} > \tau_\eta^{2(t-1)}$. The parameters $(\sigma_{b_2}^{2(t)}, \tau^{2(t)})$ can be sampled in the same way as τ_η^2, by again evaluating the likelihood with the updated τ_η^2 and drawing new ν_1 and z_1.

Next we also apply the shrinking sampling technique to $\mathbf{\Sigma}_\alpha$. However, applying it is not straightforward because $\mathbf{\Sigma}_\alpha$ has 3 parameters. Even if we employ the multivariate shrinkage sampling method described in *Neal* (2003), sampling 3 parameters of $\mathbf{\Sigma}_\alpha$ under multi-dimensional bounds is computationally inefficient. Hence, since the prior distribution of $\mathbf{\Sigma}_\alpha$ is an inverse Wishart with a diagonal scale matrix, we re-parameterized $\mathbf{\Sigma}_\alpha$ by a product of inverse gamma and normal form (*Gelfand et al.*, 2004). For example, $\mathbf{\Sigma}_\alpha \sim IW_2(\nu, \Delta^{-1})$ where Δ is a diagonal scale matrix with the i^{th} diagonal element $\Delta_i, i = 1, 2$. Let $\nu_{11} \sim IG((\nu-1)/2, \Delta_1/2), \nu_{22} \sim IG(\nu/2, \Delta_2/2), \nu_{12}|\nu_{22} \sim N(0, \nu_{22}/\Delta_1)$. Then, $\mathbf{\Sigma}_\alpha = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix} = \begin{bmatrix} \nu_{11}^2 & \nu_{12}\nu_{11} \\ \nu_{12}\nu_{11} & \nu_{12}^2\nu_{11} + \nu_{22}^2 \end{bmatrix}$. Re-parameterizing $\mathbf{V} = \begin{bmatrix} \nu_{11} & v_{12} \\ \nu_{12} & \nu_{22} \end{bmatrix}$ to $\mathbf{\Sigma}_\alpha$ can be easily done with the Sweep operator (*Beaton*, 1964; *Dempster*, 1969). The

relationship between $\boldsymbol{\Sigma}_\alpha$ and $\mathbf{V}$ is

$$RSW[2]\begin{bmatrix} & SWP[2]\begin{bmatrix} 0 & \frac{1}{r} & 0 \\ \frac{1}{r} & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} & \\ & & \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & \sigma_{11}^2 & \sigma_{12} \\ 0 & \sigma_{12} & \sigma_{22}^2 \end{bmatrix}, \tag{E.3}$$

where $SWP[j]$ and $RSW[j]$ are sweeping and reverse sweeping operations on the j^{th} row and column, respectively. Hence, instead of sampling $\boldsymbol{\Sigma}_\alpha$, we used the shrinkage slice sampling for $\nu_{ii}, i = 1, 2$, from $IG(\frac{\nu-(2-i)}{2}, \frac{\Delta_i}{2})$, and for ν_{12} from $\nu_{12}|\nu_{22} \sim N(0, \frac{\nu_{22}}{\Delta_1})$. Each element of $\mathbf{V}$ is updated at a time and then $\mathbf{V}$ immediately transformed to $\boldsymbol{\Sigma}_\alpha$ using the sweep/reverse operators.

Given posterior samples of $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\alpha}, \boldsymbol{b}_1)$ and $\boldsymbol{\Omega} = (\tau_\eta^2, \boldsymbol{\Sigma}_\alpha, \sigma_{b_2}^2, \tau^2)$, we estimate each district's intercept and DL effects by sampling district random effects $\eta_j, \boldsymbol{\alpha}_j, \boldsymbol{b}_{1j}$. Let $\boldsymbol{r}_j = \mathbf{Y}_j - \beta_0 - \mathbf{X}_j^*\boldsymbol{\alpha} - \mathbf{Z}_j^*\boldsymbol{b}_1$. To sample $(\eta_j, \boldsymbol{\alpha}_j)$, let $\boldsymbol{\Psi}_{1j} = \sigma_{b_2}^2\mathbf{Z}_j^*\mathbf{Z}_j^{*T} + \tau^2\mathbf{I}_{n_j}$. Then, $(\eta_j, \boldsymbol{\alpha}_j) \sim N(\mathbf{A}_j\mathbf{C}_j, \mathbf{A}_j)$, where $\mathbf{A}_j = \left([\mathbf{1}_{n_j}\ \mathbf{X}_j^*]^T\boldsymbol{\Psi}_{1j}^{-1}[\mathbf{1}_{n_j}\ \mathbf{X}_j^*] + \begin{bmatrix} \tau_\eta^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\alpha \end{bmatrix}\right)^{-1}$ and $\mathbf{C}_j = [\mathbf{1}_{n_j}\ \mathbf{X}_j^*]^T\boldsymbol{\Psi}_{1j}^{-1}\boldsymbol{r}_j$.

To sample $\boldsymbol{b}_{1j}$, let $\boldsymbol{\Psi}_{2j} = [\mathbf{1}_{n_j}\ \mathbf{X}_j^*]\begin{bmatrix} \tau_\eta^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\alpha \end{bmatrix}[\mathbf{1}_{n_j}\ \mathbf{X}_j^*]^T + \tau^2\mathbf{I}_{n_j}$. Then, $\boldsymbol{b}_{1j} \sim N(\mathbf{A}_j^b\mathbf{C}_j^b, \mathbf{A}_j^b)$, where $\mathbf{A}_j^b = (\mathbf{Z}_j^{*T}\boldsymbol{\Psi}_{2j}^{-1}\mathbf{Z}_j^* + \sigma_{b_2}^{-2}\mathbf{I}_{L-2})^{-1}$ and $\mathbf{C}_j^b = \mathbf{Z}_j^{*T}\boldsymbol{\Psi}_{2j}^{-1}\boldsymbol{r}_j$. With sampled $(\boldsymbol{\alpha}_j, \boldsymbol{b}_{1j})$ for a district j, we estimate district-specific DL effects

$$\boldsymbol{\beta}_j = (\beta_j(0; r_1), \beta_j(r_1; r_2), \ldots, \beta_j(r_{L-1}; r_L))^T = [\boldsymbol{C}_0\ \boldsymbol{C}_1\mathbf{M}_1\mathbf{M}_2^{-1/2}]\begin{bmatrix} \boldsymbol{\alpha} + \boldsymbol{\alpha}_j \\ \boldsymbol{b}_1 + \boldsymbol{b}_{1j} \end{bmatrix}.$$

Figure E.1: **The 7^{th} grade children's mean BMIz by Assembly districts in 2001, 2005, 2010 in a whole CA, LA and SF metropolitan areas.**

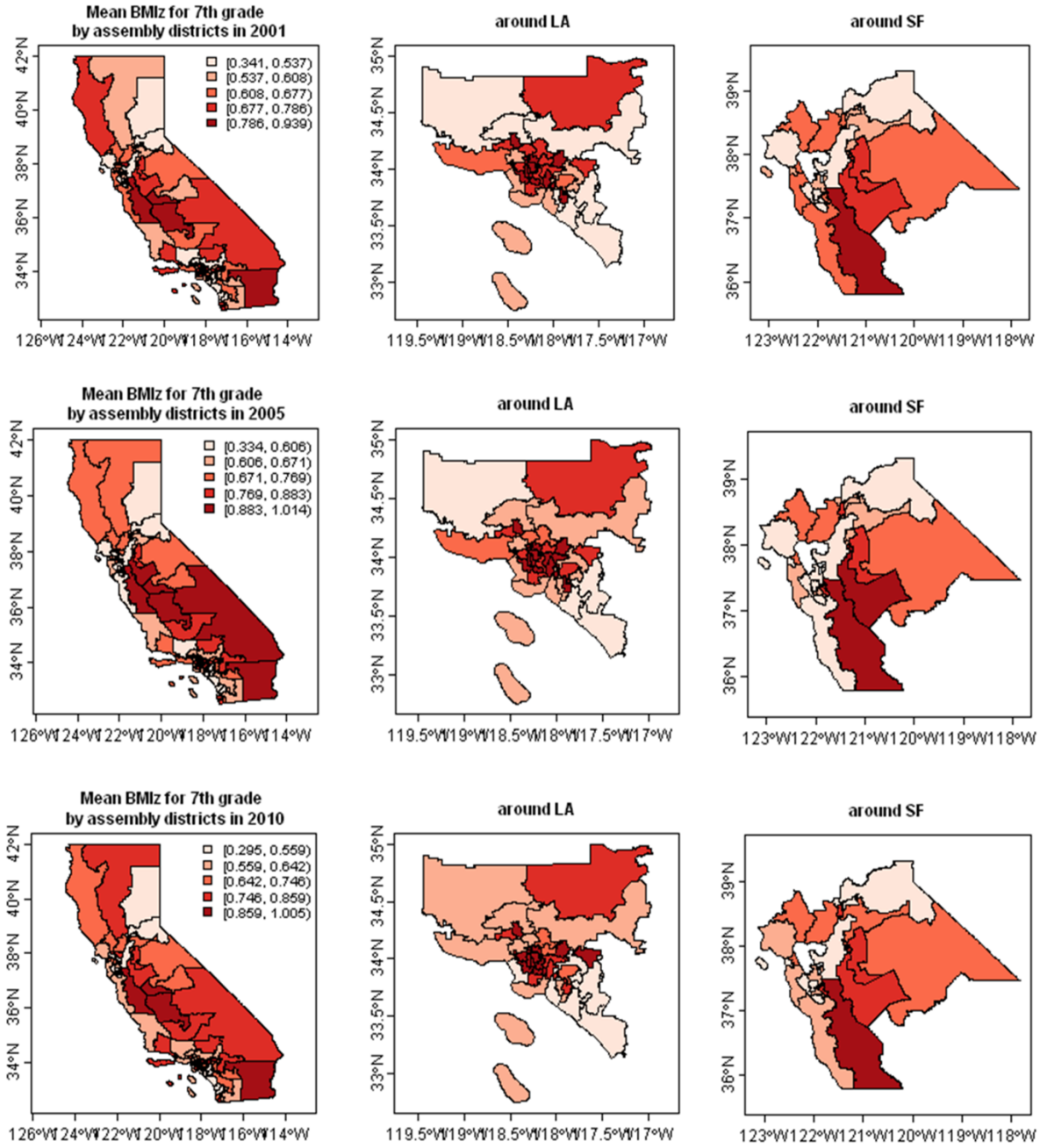


Figure E.2: **(a) Estimated DL coefficients of features of the built environment by Assembly districts. (b) Histogram of estimated DL coefficients smoothness paramters by Assembly districts. (c) Two categorized groups of Assembly districts for smoothness parameters of DL coefficients.**

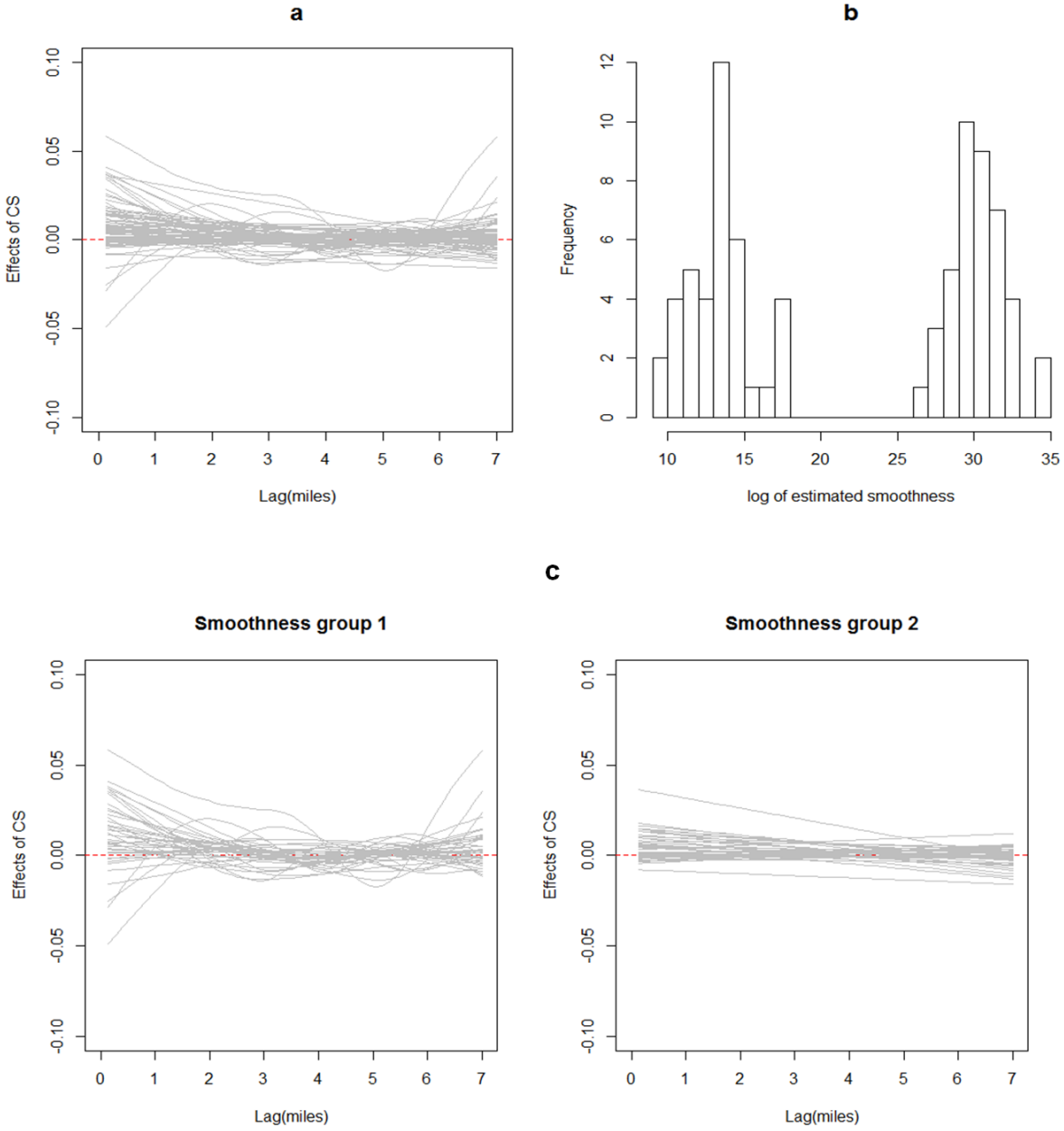


Figure E.3: **Estimated buffer effects up to (a) 1/4, (b) 1/2, (c) 3/4 miles from schools by Assembly districts in a whole CA, LA and SF metropolitan areas. The district-specific buffer effects are estimated by each subset of Assembly districts in 2001-2010**

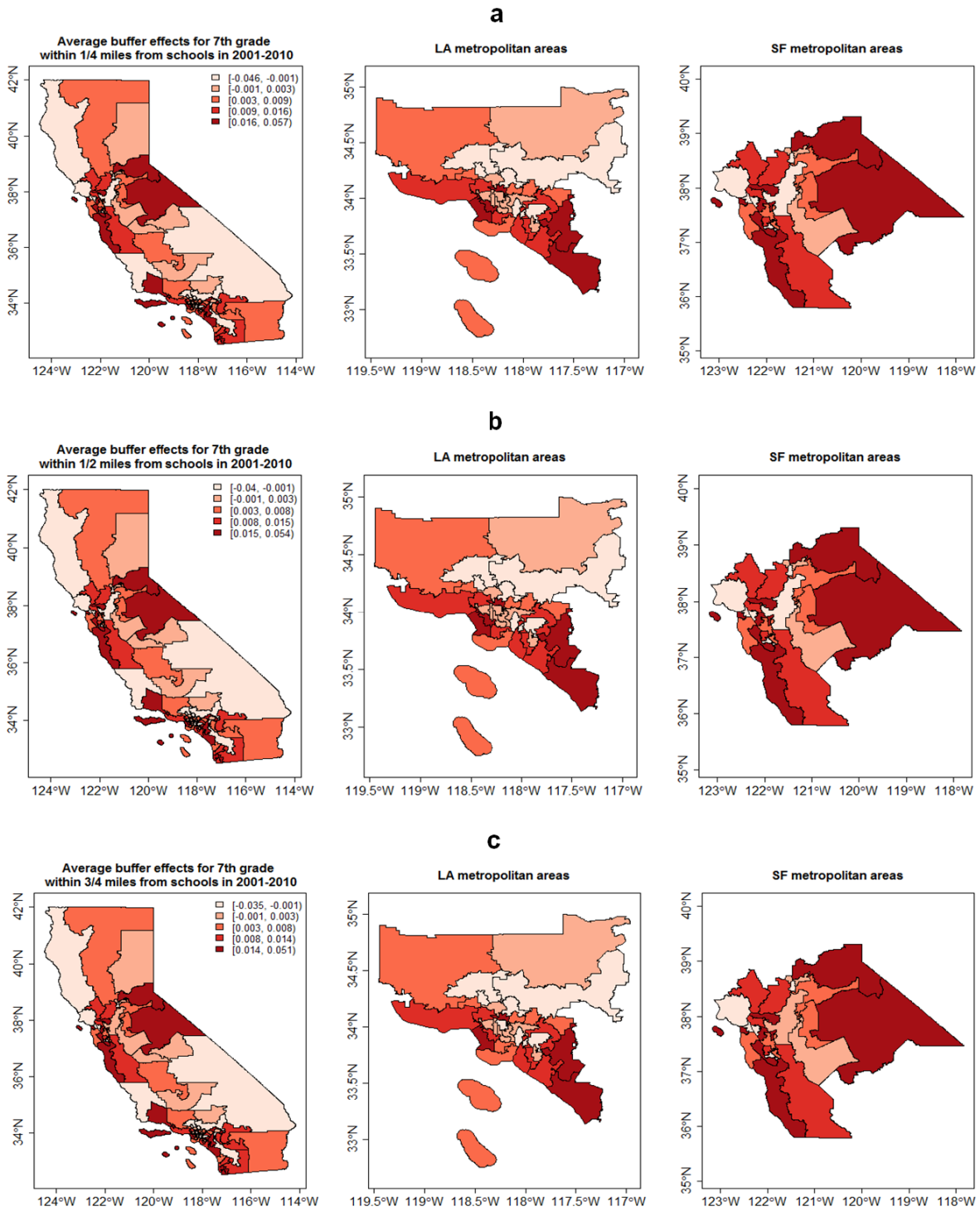


Figure E.4: **(a) The estimated overall DL effects and (b) the estimated random DL effects of Assembly districts from the individual and school characteristics adjusted HDLM (model 3), (c) the estimated mile buffer effects of Assembly districts.**

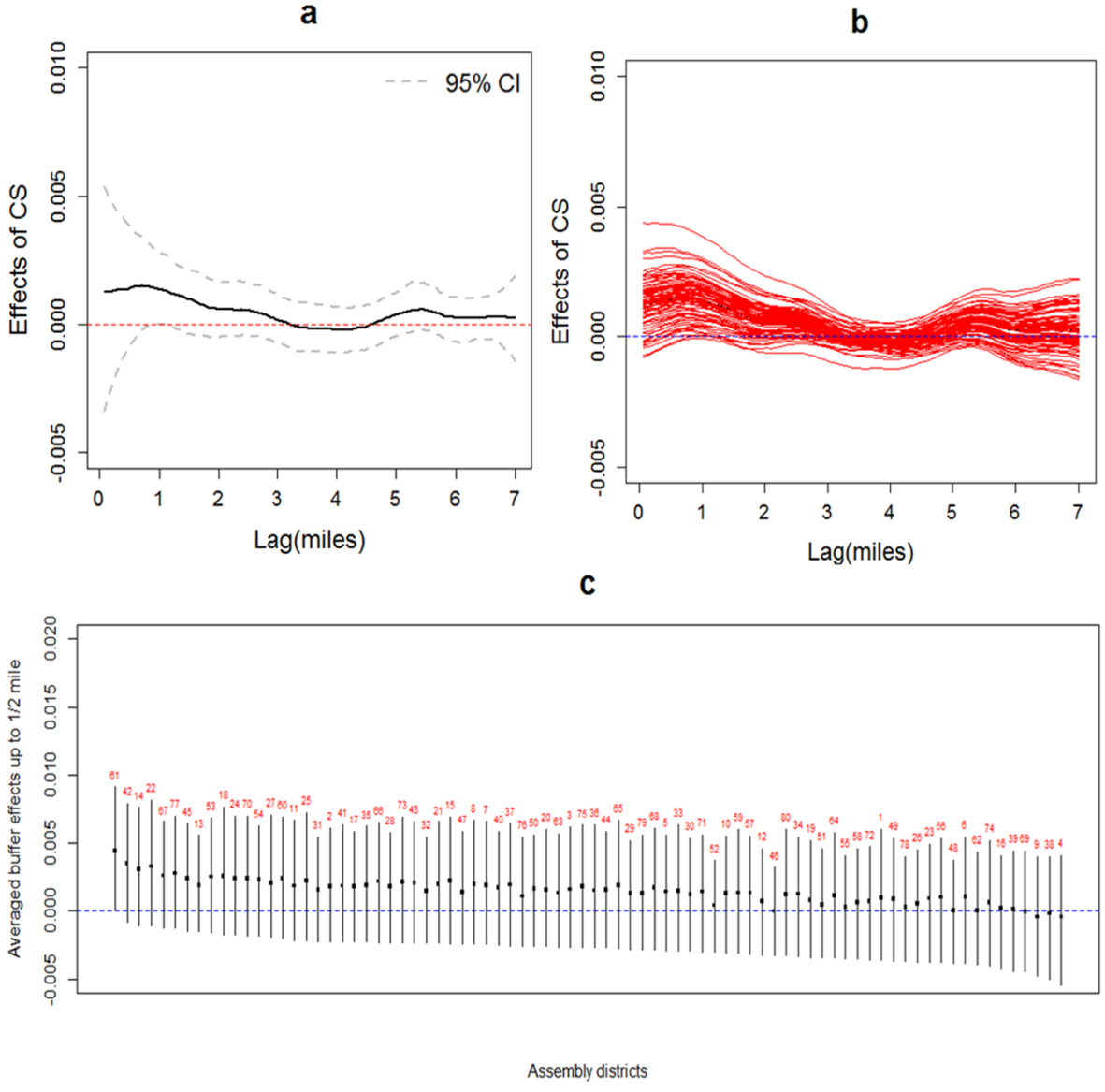
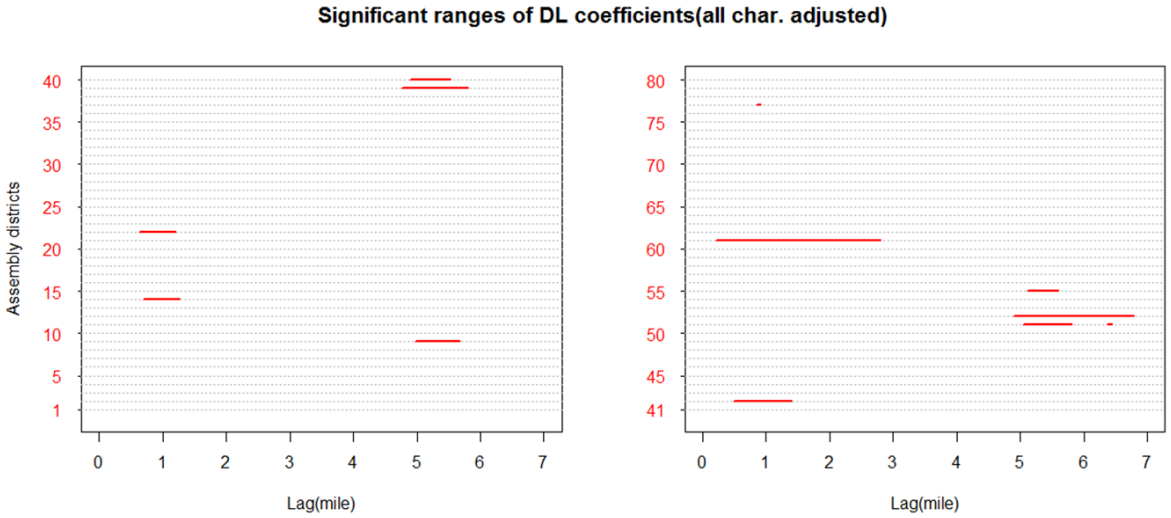
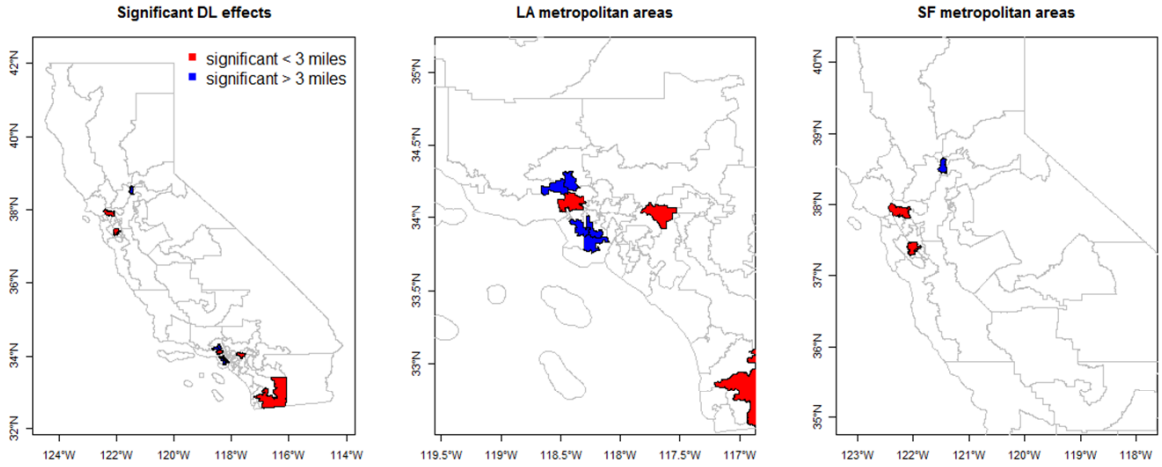


Figure E.5: **(a) Statistical significance of DL effects up to 7 miles across 80 Assembly districts. (b) Mapped Assembly districts with significant DL effects before (red) and after (blue) 3 miles. Implemented HDLMs adjusted both individual and school characteristics adjusted.**

a



b



BIBLIOGRAPHY

BIBLIOGRAPHY

Agarwal, D. K., and A. E. Gelfand (2005), Slice sampling for simulation based fitting of spatial data models, *Statistics and Computing*, *15*(1), 61–69, doi:10.1007/s11222-005-4790-z.

Almon, S. (1965), The distributed lag between capital appropriations and expenditures, *Econometrica: Journal of the Econometric Society*, pp. 178–196.

Alviola, P. a., R. M. Nayga, M. R. Thomsen, D. Danforth, and J. Smartt (2014), The effect of fast-food restaurants on childhood obesity: a school level analysis., *Economics and human biology*, *12*, 110–119, doi:10.1016/j.ehb.2013.05.001.

An, R., and R. Sturm (2012), School and residential neighborhood food environment and diet among California youth., *American journal of preventive medicine*, *42*(2), 129–135, doi:10.1016/j.amepre.2011.10.012.

and others ().

Apparicio, P., M. Abdelmajid, M. Riva, and R. Shearmur (2008), Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues., *International journal of health geographics*, *7*(7), doi:10.1186/1476-072X-7-7.

Austin, S. B., S. J. Melly, B. N. Sanchez, A. Patel, S. Buka, and S. L. Gortmaker (2005), Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments., *American journal of public health*, *95*(9), 1575–1581, doi:10.2105/AJPH.2004.056341.

Baek, J., B. N. Sánchez, and E. V. Sanchez-Vaznaugh (2014), Hierarchical multiple informants models: examining food environment contributions to the childhood obesity epidemic., *Statistics in Medicine*, *33*(4), 662–674, doi:10.1002/sim.5967.

Beaton, A. E. (1964), The use of special matrix operators in statistical calculus, *Educational Testing Service*, pp. RB–64–51.

Berhane, K., and D. C. Thomas (2002), A two-stage model for multiple time series data of counts., *Biostatistics (Oxford, England)*, *3*(1), 21–32, doi:10.1093/biostatistics/3.1.21.

Besag, J., J. York, and A. Mollii (1991), Bayesian image restoration, with two applications in spatial statistics., *Annals of the Institute of Statistical Mathematics*, *43*(1), 1–20.

Black, J. L., and J. Macinko (2008), Neighborhoods and obesity., *Nutrition reviews*, *66*(1), 2–20, doi:10.1111/j.1753-4887.2007.00001.x.

Braga, A. L. F., A. Zanobetti, and J. Schwartz (2014), The Time Course of Weather-Related Deaths, *Epidemiology*, *12*, 662–667.

CDC (2005), Centers for Disease Control and Prevention. CDC growth charts, United States, *Atlanta (GA): CDC 2005; Retrieved January 28, 2010, http://www.cdc.gov/GROWTHcharts/.*

Chaix, B., C. Leal, and D. Evans (2010), Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions., *Epidemiology*, *21*(1), 124–127, doi:10.1097/EDE.0b013e3181c04e70.

Clayton, D., and J. Kaldor (1987), Empirical Bayes estimates of age-standardized relative risks for use in disease mapping., *Biometrics*, *43*(3), 671–681.

Cummins, S., S. Curtis, A. V. Diez-Roux, and S. Macintyre (2007), Understanding and representing 'place' in health research: a relational approach., *Social science & medicine (1982)*, *65*(9), 1825–1838, doi:10.1016/j.socscimed.2007.05.036.

Currie, J., S. Dellavigna, E. Moretti, and V. Pathania (2009), The effect of fast food restaurants on obesity and weight gain, *Cambridge, MA: National Bureau of Economic Research*, pp. Working Paper No. 14,721. JEL No. I1, I18, J0.

Daniels, S. R. (2006), The Consequences of Childhood Overweight and Obesity, *The Future of Children*, *16*(1), 47–67, doi:10.1353/foc.2006.0004.

Davis, B., and C. Carpenter (2009), Proximity of fast-food restaurants to schools and adolescent obesity, *American Journal of Public Health*, *99*(3), 505–510, doi: 10.2105/AJPH.2008.137638.

Dempster, A. P. (1969), *Elements of continuous multivariate analysis*, 388 p. pp., Addison-Wesley Pub. Co. (Reading, Mass).

Diez-Roux, A. V. (1998), Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis, *American journal of public health*, *88*(2), 216–222.

Diez Roux, A. V. (2004), Estimating neighborhood health effects: the challenges of causal inference in a complex world., *Social science & medicine (1982)*, *58*(10), 1953–1960, doi:10.1016/S0277-9536(03)00414-3.

Dominici, F., J. M. Samet, S. L. Zeger, S. Journal, R. Statistical, S. Series, A. Statistics, and M. Samet (2000), Hierarchical Modelling Strategy Combining evidence on air pollution and daily mortality from the 20 largest US cities : a hierarchical

modelling strategy, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *163*(3), 263–302.

Dominici, F., A. McDermott, and T. J. Hastie (2004), Improved Semiparametric Time Series Models of Air Pollution and Mortality, *Journal of the American Statistical Association*, *99*(468), 938–948, doi:10.1198/016214504000000656.

Drewnowski, A., C. Rehm, C. Kao, and H. Goldstein (2009), Poverty and childhood overweight in California Assembly districts., *Health & place*, *15*(2), 631–635, doi:10.1016/j.healthplace.2008.09.008.

Durbán, M., J. Harezlak, M. P. Wand, and R. J. Carroll (2005), Simple fitting of subject-specific curves for longitudinal data., *Statistics in medicine*, *24*(8), 1153–1167, doi:10.1002/sim.1991.

Flowerdew, R., D. J. Manley, and C. E. Sabel (2008), Neighbourhood effects on health: does it matter where you draw the boundaries?, *Social science & medicine (1982)*, *66*(6), 1241–1255, doi:10.1016/j.socscimed.2007.11.042.

Fotheringham, a. S., and D. W. S. Wong (1991), The modifiable areal unit problem in multivariate statistical analysis, *Environment and Planning A*, *23*(7), 1025–1044, doi:10.1068/a231025.

Gasparrini, A. (2011), Distributed Lag Linear and Non-Linear Models in R: The Package dlnm., *Journal of statistical software*, *43*(8), 1–20.

Gebauer, H., and M. N. Laska (2011), Convenience stores surrounding urban schools: an assessment of healthy food availability, advertising, and product placement., *Journal of urban health : bulletin of the New York Academy of Medicine*, *88*(4), 616–622, doi:10.1007/s11524-011-9576-3.

Gelfand, A. E., H.-J. Kim, C. F. Sirmans, and S. Banerjee (2003), Spatial Modeling With Spatially Varying Coefficient Processes, *Journal of the American Statistical Association*, *98*(462), 387–396, doi:10.1198/016214503000170.

Gelfand, A. E., A. M. Schmidt, S. Banerjee, and C. F. Sirmans (2004), Nonstationary multivariate process modeling through spatially varying coregionalization, *Test*, *13*(2), 263–312, doi:10.1007/BF02595775.

Goodman, P. G., D. W. Dockery, and L. Clancy (2004), Cause-specific mortality and the extended effects of particulate pollution and temperature exposure, *Environmental Health Perspectives*, *112*(2), 179–185, doi:10.1289/ehp.6451.

Green, P. J., and B. W. Silverman (1993), *Nonparametric regression and generalized linear models: a roughness penalty approac*, CRC Press.

Guo, J., and C. Bhat (2004), Modifiable Areal Units: Problem or Perception in Modeling of Residential Location Choice?, *Transportation Research Record*, *1898*(1), 138–147, doi:10.3141/1898-17.

Harris, D. E., J. W. Blum, M. Bampton, L. M. O'Brien, C. M. Beaudoin, M. Polacsek, and K. a. O'Rourke (2011), Location of food stores near schools does not predict the weight status of Maine high school students., *Journal of nutrition education and behavior*, *43*(4), 274–278, doi:10.1016/j.jneb.2010.08.008.

Hastie, T. J., and R. J. Tibshirani (1990), *Generalized additive models*, CRC Press.

Heaton, M. J., and A. E. Gelfand (2011), Spatial Regression Using Kernel Averaged Predictors, *Journal of Agricultural, Biological, and Environmental Statistics*, *16*(2), 233–252, doi:10.1007/s13253-010-0050-6.

Heaton, M. J., and R. D. Peng (2012), Flexible Distributed Lag Models using Random Functions with Application to Estimating Mortality Displacement from Heat-Related Deaths., *Journal of agricultural, biological, and environmental statistics*, *17*(3), 313–331, doi:10.1007/s13253-012-0097-7.

Hillier, A., B. L. Cole, T. E. Smith, A. K. Yancey, J. D. Williams, S. a. Grier, and W. J. McCarthy (2009), Clustering of unhealthy outdoor advertisements around child-serving institutions: a comparison of three cities., *Health & place*, *15*(4), 935–945, doi:10.1016/j.healthplace.2009.02.014.

Horton, N. J., and G. M. Fitzmaurice (2004), Regression analysis of multiple source and multiple informant data from complex survey samples., *Statistics in Medicine*, *23*(18), 2911–2933, doi:10.1002/sim.1879.

Horton, N. J., N. M. Laird, and G. E. Zahner (1999), Use of multiple informant data as a predictor in psychiatric epidemiology, *International Journal of Methods in Psychiatric Research*, *8*(1), 6–18, doi:10.1002/mpr.52.

Howard, P. H., M. Fitzpatrick, and B. Fulfrost (2011), Proximity of food retailers to schools and rates of overweight ninth grade students: an ecological study in California., *BMC public health*, *11*(1), 68, doi:10.1186/1471-2458-11-68.

Huang, Y., F. Dominici, and M. L. Bell (2005), Bayesian hierarchical distributed lag models for summer ozone exposure and cardio-respiratory mortality., *Environmetrics*, *16*(5), 547–562, doi:10.1002/env.721.

Kipke, M. D., E. Iverson, D. Moore, C. Booker, V. Ruelas, A. L. Peters, and F. Kaufman (2007), Food and park environments: neighborhood-level risks for childhood obesity in east Los Angeles., *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, *40*(4), 325–333, doi:10.1016/j.jadohealth.2006.10.021.

Koyck, L. M. (1954), *Distributed lags and investment analysis*, Amsterdam: North-Holland Publishing Company.

Kwan, M.-P. (2012), The Uncertain Geographic Context Problem, *Annals of the Association of American Geographers*, *102*(5), 958–968, doi:10.1080/00045608.2012.687349.

Langellier, B. a. (2012), The food environment and student weight status, Los Angeles County, 2008-2009., *Preventing chronic disease*, *9*(9), E61.

Leal, C., K. Bean, F. Thomas, and B. Chaix (2011), Are associations between neighborhood socioeconomic characteristics and body mass index or waist circumference based on model extrapolations?, *Epidemiology (Cambridge, Mass.)*, *22*(5), 694–703, doi:10.1097/EDE.0b013e3182257784.

Liang, K. Y., and S. L. Zeger (1986), Longitudinal data analysis using generalized linear models, *Biometrika*, *73*(1), 13–22.

Litman, H. J., N. J. Horton, B. Hernández, and N. M. Laird (2008), Estimation of Marginal Regression Models with Multiple Source Predictors, *Handbook of Statistics*, *27*(07), 730–746, doi:10.1016/S0169-7161(07)27025-7.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000), WinBUGS A Bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing*, *10*(4), 325–337.

Macnab, Y. C., and P. Gustafson (2007), Regression B-spline smoothing in Bayesian disease mapping : With an application to patient safety surveillance, *Statistics in Medicine*, *26*, 4455–4474, doi:10.1002/sim.

Madden, L. V., and P. a. Paul (2010), An assessment of mixed-modeling approaches for characterizing profiles of time-varying response and predictor variables., *Phytopathology*, *100*(10), 1015–1029, doi:10.1094/PHYTO-01-10-0001.

Mancl, L. A., and T. A. Derouen (2001), A covariance estimator for GEE with improved small-sample properties, *Biometrics*, *57*, 126–134.

Mancl, L. A., and B. G. Leroux (1996), Efficiency of regression estimates for clustered data., *Biometrics*, *52*(2), 500–511.

Must, A., and S. E. Anderson (2006), Body mass index in children and adolescents: considerations for population-based applications., *International journal of obesity*, *30*(4), 590–594, doi:10.1038/sj.ijo.0803300.

Neal, R. M. . (2003), Slice Sampling, *The Annals of Statistics*, *31*(3), 705–741.

Nerlove, M. (1956), Estimates of the elasticities of supply of selected agricultural commodities, *Journal of Farm Economics*, *38*(2), 496–509.

N.P. (2013), California State Assembly (http://assembly.ca.gov/sites/assembly.ca.gov/files/2013-14_CaliforniaAsmPamphlet.pdf).

N.P. (2014), THE FOOD LANDSCAPE IN CALIFORNIA CITIES AND COUNTIES (http://publichealthadvocacy.org/searchingforhealth).

Openshaw, S. (1984), *The Modifiable Areal Unit Problem (Concepts and Techniques in Modern Geography)*., Geo Books.

Openshaw, S. (1996), *Spatial Analysis: Modelling in a GIS Environment. Chapter 4. Developing GIS-relevant zone-based spatial analysis methods.*, 55–73 pp.

Pan, W., T. A. Louis, J. E. Connett, and E. Connett (2000), A Note on Marginal Linear Regression With Correlated Response Data, *The American Statistician*, *54*(3), 191–195.

Pearce, J., K. Witten, and P. Bartie (2006), Neighbourhoods and health: a GIS approach to measuring community resource accessibility., *Journal of epidemiology and community health*, *60*(9), 389–395, doi:10.1136/jech.2005.043281.

Peng, R. D., F. Dominici, L. J. Welty, S. Journal, R. Statistical, S. Series, and C. A. Statistics (2009), A Bayesian Hierarchical Distributed Lag Model for Estimating the Time Course of Risk of Hospitalization Associated with Particulate Matter Air Pollution estimating associated distributed hierarchical lag model for the time course of risk of hospitalizatio, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *58*(1), 3–24.

Pepe, M. S., and G. L. Anderson (1994), A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data, *Communications in Statistics*, *23*(4), 939–951, doi:10.1080/03610919408813210.

Pepe, M. S., R. C. Whitaker, and K. Seidel (1999), Estimating and comparing univariate associations with application to the prediction of adult obesity, *Statistics in Medicine*, *18*(2), 163–173, doi:DOI: 10.1002/(SICI)1097-0258(19990130)18:2¡163::AID-SIM11¿3.0.CO;2-F.

Pope, C. A., and J. Schwartz (1996), Time series for the analysis of pulmonary health data., *American journal of respiratory and critical care medicine*, *154*, S229–S233.

Pope, C. A., D. W. Dockery, J. D. Spengler, and M. E. Raizenne (1991), Respiratory Health and PM 10 Pollution, *American Review of Respiratory Diseases*, *144*, 668–674.

R Development Core Team (2014), R: a language and environment for statistical computing, reference index version 3.0.1 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http::/www.R-project.org.

Rahman, T., R. A. Cushing, and R. J. Jackson (2011), Contributions of Built Environment to Childhood Obesity., *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, *78*(1), 49–57, doi:10.1002/MSJ.

Reilly, J. J., and J. Kelly (2011), Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review., *International journal of obesity (2005)*, *35*(7), 891–898, doi:10.1038/ijo.2010.222.

Reilly, J. J., E. Methven, Z. C. McDowell, B. Hacking, D. Alexander, L. Stewart, and C. J. H. Kelnar (2003), Health consequences of obesity., *Archives of disease in childhood*, *88*(9), 748–752.

Rochon, J. (1996), Analyzing bivariate repeated measures for discrete and continuous outcome variables, *Biometrics*, *52*(2), 740–750.

Rondeau, V., K. Berhane, and D. C. Thomas (2005), A three-level model for binary time-series data: the effects of air pollution on school absences in the Southern California Children's Health Study., *Statistics in medicine*, *24*(7), 1103–1115, doi: 10.1002/sim.1980.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003), *Semiparametric regression*, Cambridge University Press.

Sallis, J. F., and K. Glanz (2006), The Role of Built Environments in Physical Activity, Eating, and Obesity in Childhood, *The Future of Children*, *16*(1), 89–108, doi:10.1353/foc.2006.0009.

Sánchez, B. N., E. V. Sanchez-Vaznaugh, A. Uscilka, J. Baek, and L. Zhang (2012), Differential associations between the food environment near schools and childhood overweight across race/ethnicity, gender, and grade., *American journal of epidemiology*, *175*(12), 1284–1293.

Sanchez-Vaznaugh, E. V., B. N. Sánchez, J. Baek, and P. B. Crawford (2010), 'Competitive' food and beverage policies: are they influencing childhood overweight trends?, *Health affairs (Project Hope)*, *29*(3), 436–446, doi: 10.1377/hlthaff.2009.0745.

Schaefer-McDaniel, N., M. O. Caughy, P. O'Campo, and W. Gearey (2010), Examining methodological details of neighbourhood observations and the relationship to health: a literature review., *Social science & medicine*, *70*(2), 277–292, doi: 10.1016/j.socscimed.2009.10.018.

Singh, G. K., M. Siahpush, and M. D. Kogan (2010), Neighborhood socioeconomic conditions, built environments, and childhood obesity., *Health affairs (Project Hope)*, *29*(3), 503–512, doi:10.1377/hlthaff.2009.0730.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. V. D. Linde (2002), Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society. Series B, Statistical methodology*, *64*(4), 583–639.

Spielman, S. E., and E.-H. Yoo (2009), The spatial dimensions of neighborhood effects., *Social science & medicine*, *68*(6), 1098–1105, doi: 10.1016/j.socscimed.2008.12.048.

Susser, M. (1994), The logic in ecological: I. The logic of analysis., *American journal of public health*, *84*(5), 825–829.

Thornton, L. E., J. R. Pearce, and A. M. Kavanagh (2011), Using Geographic Information Systems (GIS) to assess the role of the built environment in influencing obesity: a glossary., *The international journal of behavioral nutrition and physical activity*, *8*(10), doi:10.1186/1479-5868-8-71.

Vallée, J., and P. Chauvin (2012), Investigating the effects of medical density on health-seeking behaviours using a multiscale approach to residential and activity spaces: results from a prospective cohort study in the Paris metropolitan area, France., *International journal of health geographics*, *11*, 54, doi:10.1186/1476-072X-11-54.

Vallée, J., G. Le Roux, B. Chaix, Y. Kestens, and P. Chauvin (2014), The 'constant size neighbourhood trap' in accessibility and health studies, *Urban Studies*, pp. 1–20, doi:10.1177/0042098014528393.

Welty, L. J., R. D. Peng, S. L. Zeger, and F. Dominici (2009), Bayesian distributed lag models: estimating effects of particulate matter air pollution on daily mortality., *Biometrics*, *65*(1), 282–291, doi:10.1111/j.1541-0420.2007.01039.x.

Zanobetti, A., M. P. Wand, J. Schwartz, and L. M. Ryan (2000), Generalized additive distributed lag models : quantifying, *Biostatistics*, *1*(3), 279–292.

Zellner, A. (1962), An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, *Journal of the American Statistical Association*, *57*(298), 348–368.

Zenk, S. N., and L. M. Powell (2008), US secondary schools and food outlets., *Health & place*, *14*(2), 336–346, doi:10.1016/j.healthplace.2007.08.003.

Zhao, X., F. Chen, Z. Feng, X. Li, and X.-H. Zhou (2014), The temporal lagged association between meteorological factors and malaria in 30 counties in south-west China: a multilevel distributed lag non-linear analysis., *Malaria journal*, *13*(57), 1–12, doi:10.1186/1475-2875-13-57.