ISSN: 2641-6336
DOI: 10.33552/ABBA.2020.04.000585
Annals of
Biostatistics & Biometric Applications
Iris Publishers

Research Article

Moran Spatial Filter Eigenvector Mapping and Field Verification of Latent Non-Zero Autocorrelation Georeferenced Clusters Stratified by Homeless Time Series Socioeconomic Causation Covariates in Tampa-Hillsborough County, Florida

Christopher Villatte[1], Russel Kirby[1], Leisa Stanley[2], Ricardo Izurieta[1] and Benjamin G Jacob[1]*

[1]*College of + Public Health, University of South Florida, USA*

[2]*Healthy Start Coalition of Hillsborough Count, University of South Florida, USA*

***Corresponding author:** Benjamin G. Jacob, College of Public Health, University of South Florida, United States.

Received Date: March 2, 2021

Published Date: March 30, 2021

Abstract

In the context of spatial regression analysis, several methods are employable to control for non-asymptotic approximation effects rendered from inconspicuous spatial dependencies amongst georeferenced, homeless-related, socioeconomic stratified, time series, observational prognosticators. Maximum likelihood or Bayesian approaches account for spatial dependencies in a parametric framework, whereas recent spatial filtering approaches focus on non-parametrically removing spatial autocorrelation. In this article we propose a semiparametric spatial filtering approach that allows homeless researchers to deal explicitly with (a) spatially lagged autoregressive models and (b) simultaneous autoregressive spatial models. Our primary assumption was temporally dependent, homeless stratified, socioeconomic, causation covariate, clustering propensities may be revealed employing orthogonal, synthetic, eigenfunction, spatial filters. We created a spatial weights matrix in PROC AUTOREG so that neighboring socioeconomic stratified, frequency samples received a weight that was proportional to the calculable inverse distance measurement between a geographic sub-county, time series, sampled geographic location and its neighbor. We spatially tabularized Euclidean distances in ArcGIS along the links in the eigenfunction eigen decomposition analysis. Our hypothesis was that multivariate autoregressively dependent, diagnostic, frequency model, spatial filter eigenvectors could cartographically and geo-statistically distinguish among the effects of non-parameterizable non-Gaussian non-normalities [e.g., spatial heteroscedasticity(i.e, uncommon variance)],in Euclidean distance measurements between dereferenceable homeless geographically stratified (henceforth geo-stratified), hot and cold spot, sub-county clusters employing a stochastic simulation of temporally regressable, socioeconomic indexed prognosticators. As in one nonparametric spatial filtering approach, a specific subset of eigenvectors from the transformed spatial link matrix in PROC AUTOREG captured dependencies among the disturbances in the empirically stratified datasets of the regressed, homeless, socioeconomic cluster model eigen-estimators. However, the optimal subset in the proposed orthogonal spatial filtering model was identified more intuitively by an objective function that minimized latent, non-zero autocorrelation, in the sampled eigenvectors rather than maximized a model fit. The proposed objective function had the advantage that it lead to a robust and smaller subset of parsimoniously selected eigenvectors. An application of the proposed eigenvector spatial filtering approach in Proc Autoreg employed an empirical parameter estimator dataset for optimally delineating the sub-county georeferenced hot/cold spot clusters in Tampa-Hillsborough County. The top causes of homelessness based on the eigenvector, spatial filter geo-stratified, high positively autocorrelated, georeferenceable, hot spot clusters were (1) unemployment and 2) drug usage/ transaction. In slightly positive autocorrelated clusters homelessness causation was identified as 1) previous incarceration, 2) medical care/ food shortage and 3) domestic violence especially for female victims. Mental health was the primary, diagnostic frequency covariate in the residually negatively autocorrelated clusters. The vast majority interviewed in the negatively autocorrelated cluster during field validation ("ground trothing") exercises had severe psychological illnesses that remained largely untreated. An observational study found significant levels of mobility aid are required among the homeless in Tampa. A collection of time series, spatial autocorrelation socioeconomic, time sensitive, stratified, frequency, cluster indexed maps should be created and field validated. These maps may be employable by health and human service agencies in Tampa-Hillsborough County to predictively count the population, inventory resources, and increase awareness of targeted services especially for homeless pregnant women and children using ArcGIS and SAS predictive analytical tools. These time series autocorrelation maps may be constructed employing a selection of eigen decomposable eigen-orthogonalize, synthetic eigenvectors as rendered from an empirical geographically sampled, frequency dataset of geo-stratifiable, socioeconomic cluster, causation eigen-covariates quantitated within an autocorrelation connectivity matrix in PROC AUTOREG. New supportive facilities and shelters for the homeless should be located in areas with a high availability of employment, inexpensive or free medical care and food in Tampa-Hillsborough County. Furthermore, free mobile drug addiction programs and family domestic violence interventions should be implemented in the county. To diagnose residual autocorrelation, in empirically sampled, time series, homeless, socioeconomic, diagnostic covariates, Proc Autoreg procedure can perform a first order autocorrelation employing generalized Durbin-Watson (DW) statistics and their marginal probabilities. Exact p-values may be reported for generalized DW tests to any specified order in an homeless socioeconomic, time series, cluster model. Constructing a Bayesian Hierarchical Clustering (BHC) algorithm in Python may efficiently reveal clustering georeferenced stratified socioeconomic, time series covariates. This algorithm may define a probabilistic model of sub-county, homeless socioeconomic datasets which may be used to compute the predictive distribution of a sampled socioeconomic georeferenced, geo-stratified, sub-county, capture point and the probability of it belonging to any of existing clusters in the tree. The algorithm uses a model-based criterion to decide on merging clusters rather than an ad-hoc distance metric. Hence, Bayesian hypothesis testing may be used to decide which merges are advantageous and to output the recommended depth of the tree which may be interpreted as a novel fast bottom-up approximate inference method for a Dirichlet process (i.e., countably infinite) mixture, homeless socioeconomic, aggregation bias model. In so doing the BHS may allow a hierarchical representation of the sampled homeless socioeconomic data, incorporating both finer to coarser grained clusters, in such a way that a researcher can also make forecasts about new sub-county data points, compare different hierarchies in a principled manner, and automatically discover interesting levels of the hierarchy to examine.

Keywords: Autocorrelation; Eigenvector; Homeless; Semi-parametric; Spatial filters; Tamp-Hillsborough

Introduction

Historically, epidemiological, homelessness, forecast, risk models have been used to assess the likelihood of facing homelessness at the individual or household level. Most commonly, researchers have attempted to identify the socioeconomic co-factors that correspond to the spatial distribution of homelessness, employing geo referenceable time series data on intercity homelessness rates as the dependent variable. Found that the availability of low-income housing and lower per capita expenditures on mental health care were significantly related to homelessness rates but that poverty and unemployment rates were not. In a test of several more carefully specified frequency models of intercity homelessness rates, Burt (1992) found that per capita income, the poverty rate, and the proportion of single person households combined to explained more than half the variation in homelessness rates in high-growth cities. The authors interpreted this as evidence that more affluent households and a greater number of households with single people put pressure on the housing choices of poorer people. Based on these literature contributions, homelessness appears to vary by socioeconomic conditions, although specific study's findings have been inconsistent.

Longitudinal research has suggested the potential relevance of a structural and dynamic model of homelessness, and has raised questions about the adequacy of socioeconomic, indexed, point prevalence, temporal data for measuring the homelessness problem. Analyses of administrative data a national telephone survey and a housing survey in New York City found that as much as 3 percent of the population experienced an episode of "literal" homelessness between 1988 and 1992, suggesting a high degree of turnover in the homeless population. Longitudinal research based on tracked samples of homeless persons has also documented the often transitory, intermittent nature of homelessness. Most shelter users appear to mobilize resources and community ties to avoid the shelters most of the time. Hopper has characterized these informal networks as the "economies of makeshift." these support systems, and the factors that strain or enhance their supportive capacity, yet statistical and cartographic variables associated within these network paradigms are not well understood (see related discussions in Burt. An observational study was conducted in the field that found significant usage of mobility aid amongst the homeless population in Tampa. The study found that 12% of the homeless population in Tampa required mobility aid of some sort.Newer models developed by Bramley is a complex version of a simulation models for mapping homelessness in Tampa –Hillsborough County. It is the main tool currently used to produce projections of aggregate levels of homelessness across UK regions. The basic idea behind this approach is that housing needs are the outcomes of households, individuals and firms interacting in interconnected housing and labor markets. The model is based on a set of core functions employed to quantify responses to changes in economic and policy variables with respect to outcomes such as migration and household formation that can potentially determine housing needs.

Poverty unemployment, and lack of affordable housing are commonly recognized causes of homelessness in Tampa Hillsborough County (www.samhsa.govhomelessness-programs). These risk factors may be exacerbated by personal vulnerabilities such as mental and substance use disorders, trauma and violence, domestic violence, justice-system involvement, sudden serious illness, divorce, death of a partner, and disabilities. In order to implement county-level social programs in Tampa-Hillsborough County to reduce factors associated with homelessness, exact locations may need to be determined using spatial statistical algorithmic techniques. Statistical geography is the study and practice of collecting, analyzing and presenting data that has a geographic or real dimension such as census or demographics data. It uses techniques from spatial analysis, but also encompasses geographical activities such as the defining and naming of geographical regions for spatial statistical purposes.

Spatial statistical autocorrelation analysis frequently employs model-based inference, the dependability of which is based upon the correctness of posited assumptions about a model's error term. Spatial autocorrelation is the correlation among sampled frequency values of a single variable across a two-dimensional (2-D) surface that are geographically tied together by an underlying spatial structure, introducing a violation of the independent observations assumption of classical statistics. Its many interprets include: a nuisance parameter, self-correlation, map pattern, a diagnostic tool, a missing variables surrogate, redundant information, a spatial process mechanism, a spatial spillover, and the outcome of areal unit demarcation. It can be quantified with various indices, including the Geary Ratio and the Moran Coefficient, the statistically most powerful of the two measures. These two indices are negatively related (Griffith [1]).

Eigenvector spatial filter [ESF] regression incorporates spatial autocorrelation into the linear predictor of a generalized linear model (GLM) with a set of spatially structured, synthetic control variables from a spectral decomposition of the spatial connectivity matrix. If yi is a normally distributed set of spatially autocorrelated, socioeconomic stratified, homeless, time series, frequency-dependent observations taken at geolocations i=1, 2..., n, then following Griffith(2011,Theorem1-a). We may view yi as a mixture of normal distributions yi =y∗ i +θi where y∗ i ∼N (μi, σ2) is a spatially unstructured (exchangeable) process and θi ∼N (0, ω2) is a spatially structured (ordered) process. Subsequently, y|μ, ν ∼N (μ, ν2I), ν2 = σ2 +ω2 (Eqn.1.1). This conceptualization highlights how spatial autocorrelation could inflate observed variance statistically in a homeless, frequency, epidemiological, vulnerability-oriented, predictive, temporal model and, while the expectation of the mean is unaffected by spatial autocorrelation, failure to model spatial structure in any finite sample tends rendered may produce larger estimation errors.

To estimate θi and recover y∗ in a time series homeless cluster forecast, vulnerability model in this research, we employed ESF regression and generated the eigenfunctions of a transformed spatial connectivity matrix C, where cij =1 if polygons i and j were sampled neighbors (e.g., geographically sampled socioeconomic predictors) and all cii = 0. With projection matrix M = (I − 11′/n), where I was the identity matrix and 1 an n-by-one vector of ones, an eigenfunction decomposition of the Moran's Coefficient (MC) matrix (which appears in the numerator of the MC) was generated in AUTOREG which eventually rendered n mutually orthogonal, zero mean eigenvectors E from a time series dataset of socioeconomic explanators and their associated eigenvalues Λ. Each Ei delineating a socioeconomic frequency variable temporal sampled in the Tampa-Hillsborough County study site expressed a distinct georeferenced pattern of potential spatial autocorrelation, with the degree of it indexed by its corresponding eigenvalue λi.

The Moran eigenvector approach in ArcGIS involved delineating spatial patterns generated from an empirical sampled, frequency dataset of georeferenced, socioeconomic, homeless stratified, parameterized, estimator, time series, maps of spatial filter eigenvectors by choosing suitable orthogonal patterns and adding them to a generalized linear model (GLM). The Moran'*I* statistic for spatial autocorrelation was given by $I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2}$

where z_i was the deviation of a sampled, stratified, socioeconomic frequency homeless –related $_{\text{covariate}}$ attribute feature *i* from its mean (x_i-x) W_{IJ} was the spatial weight between features *i* and *j*, *n* was equal to the number of sampled georeferenced homeless explanatory features and s_0 was the aggregate of all the georeferenced weights $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}$. The z_i-score for the statistic was parsimoniously computed as $z_I = \frac{I - E[I]}{\sqrt{V[I]}}$ when E[I] = −1/ (n − 1) and I [I] = E [I^2] − E [I]^2 In practice, strong priori information on spatial autocorrelation and spatial filters is used to reduce the dimensionality of βE in predictive epidemiological models [Griffith 2005, Jacob et al. 2009]. A priori probability refers to the likelihood of an event (e.g., homelessness) occurring when there is a finite amount of outcomes (e.g., loss of employment, alcoholism, depression) and each is equally likely to occur. The outcomes in a priori probability in a time series, homeless frequency, socio economic, forecast, vulnerability model may not be influenced by the prior outcome. Or, put another way, any results to date will not give you an edge in predicting future results [e.g., hospitalization]. A priori probability is also referred to as classical probability. A priori probability in a homeless regression model may stipulate that the outcome of the next event is not contingent on the outcome of the previous event. A priori may also remove independent users of experience. Since the results are random and non-contingent, a homeless researcher or epidemiologist may not deduce the next outcome as rendered from the regression of the frequency county, predictive risk model, socioeconomic parameterizable estimators.

Since most model applications in epidemiological homeless research in the literature are concerned with positive spatial autocorrelation, (e.g., the geographic distribution of some variable across a map, high values tend to be geographic neighbors of high values, intermediate values tend to be geographic neighbors of intermediate values, and low values tend to be geographic neighbors of low values). the eigenvectors representing negative spatial autocorrelation (whose eigenvalues are negative) have not been considered. A homeless data analyst or epidemiological collaborator may be able to optimally quantitate negative spatial autocorrelation in a homeless frequency, temporal, socioeconomic, risk map by modelling coefficients set to zero. Quantitating negative spatial autocorrelation in a homeless, stratifiable, frequency, socioeconomic, prognosticative, epidemiological, risk model may be even simplified further by the exclusion of eigenvectors that represent only trace amounts of spatial autocorrelation. This may be performed by dropping all eigenvectors for which λi/λmax < T, with threshold T set at or below 0.25. In so doing, county geographic locations (henceforth, geolocations), eigen decomposable eigenvectors with their coefficients may be robustly parsimoniously estimated.

The most common eigenfuction eigen-decomposition algorithmic estimation procedure takes frequency eigenvectors from a candidate set and applies a stepwise variable selection procedure to identify a final subset of eigenvectors to include and accept the maximum likelihood estimate of their coefficients. The stepwise selection procedures. In so doing, information criteria [e.g., Akaike's Information Criteria. (AIC) or Bayesian Information Criteria (BIC), coefficient *p*-values, residual sum of squares or residual spatial autocorrelation as their objective function.

AIC and BIC are both penalized-likelihood criteria. Penalized likelihood estimation is a way to take into account model complexity when estimating parameters of different models. Basically, instead of conducting simple maximum likelihood estimation, a homeless researcher would maximize the log-likelihood minus a penalty term which, generally increases with the number of parameters. For instance, if a researcher is fitting a Gaussian mixture temporally dependent, homeless model, optimizing a penalized log-likelihood can help choose between models with a different number of mixture components, or between a model where the covariances are proportional to identity (e.g., one socioeconomic parameter per covariance) vs diagonal (dparameters) vs general positive symmetric matrices (d(d+1)/2 parameters). The intuition behind this is that adding socioeconomic parameters to a homeless model will give a better fit to the data, thus a higher like lihood. The AIC or BIC for a model is usually written in the form [$-2\log L + kp$], where L is the likelihood function, p is the number of parameters in the model, and k is 2 for AIC and $\log(n)$ for BIC.

The formulated AIC would be an estimate of a constant plus the relative distance between the unknown true likelihood function of the homeless time series data and the fitted likelihood function of the model, so that a lower AIC means a model is considered to be closer to the truth. BIC is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is considered to be more likely to be the true model. Both criteria are maybe vital for homeless forecast mapping vulnerable geographic locations in a county or district based on various assumptions and asymptotic approximations.

Alternatively argue that a fixed number of orthogonal spatial filter eigenvectors may be included based only on tessellation size. Outside of the variable selection probabilistic paradigm apply residual maximum likelihood (REML)to estimate the candidate eigenvector coefficients as a set of random effects (RE-ESF). In statistics, the restricted (or residual, or reduced) maximum likelihood (REML) approach is a particular form of maximum likelihood estimation that does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data.

Our assumption was that a stratified dataset of homeless, time series, indexed, socio-economic regression variables, temporally sampled, in Tampa Hillsborough County may be measurable at a geographic resolution and as such, they would render cluster indexed by moderate, positive and negative spatial autocorrelation. Satellite remotely sensed images tend to have strong to marked positive spatial autocorrelation. To date, few empirical examples of spatial autocorrelation in the literature have been reported on any frequency clustered, time series, stratified, socioeconomic parameter estimator homeless frequenct dataset, although it relates to situations of spatial competition which may be applicable for forecast mapping, for example, drug usage/transaction, geographic geolocations where many homeless may reside close study by Didenko and Pankratz indicated that two-thirds of people living on the streets blamed alcohol and/or drugs for their homelessness.

One of our principal assumptions about the latent autocorrelation in the time series, stratified frequency sampled, georefernced parameter estimators for mapping Tampa Hillsborough County homelessness, study site was that it was based on the concept that individual error terms originated from specific sampled, socioeconomic variables (e.g., previous incarceration of a migrant farm worker) whose entries were thoroughly mixed through randomness in regression space with multiple other diagnostic covariates. Moreover, we assumed that the probability of a homeless sampled, income–based, aggregation-related explanatory predictor, for example, taken on by one of a model's frequency, error term entries may not affect the probability of a value taken on by any of the remaining estimator error term entries (i.e., the independent observations assumed in classical statistics).Non-zero spatial autocorrelation in dereferenceable, empirical sampled, frequency, homeless, time series, county sampled, socioeconomic, cluster-stratified data would violate this assumption.

Another assumption we had was that time series, sampled, homelessness, frequency -oriented, explanatory, aggregation-biased, predictor variables without residual autocorrelation, would not exhibit a geographic expression when mapped; with it most of the explanators (e.g., georeferenced unemployment regressor) would exhibit some type of spatial organization across geographic space (e.g., geolocation of a zip code hotspot). Zero spatial autocorrelation means geographically random phenomena and chaotic landscapes Griffith [1]. In this research, spatial autocorrelation geographical methods were employed to study aspects of the homelessness in Tampa-Hillsborough County Florida.by identifying georeferenced, socioeconomic stratified, georeferenced, county clusters. We describe the form, direction and strength of the relationship exhibited by sample, quantitative, explanatory, regressively autocorrelated, temporal sampled, homelessness, stratified, cluster, frequency variables measured by a single set of *n* socioeconomic observations. A scatterplot in ArcGIS is employed to visualize this relationship, with a conventional correlation coefficient describing the direction and strength of a straight-line relationship of the overall homeless-related, occurrence, abundance and distribution, and their temporal frequency patterns. A variant of conventional correlation [i.e., serial correlation] pertaining to the correlation between sampled, stratified, homelessness, frequency-oriented endogenous prognosticators such as presence of mental illness and others were identified.

Morans *I* coefficient estimates were constructed which quantitated the relationship between the sampled, explanatory socioeconomic stratified, discrete intger values in the frequency-oriented, sampled dataset of homeless explanatory georeferenced variables at one geolocation (e.g., presence of food refugee camp) in geographic space and subsequently cartographically quantitated nearby sampled values in ArcGIS [e.g., Euclidean distance measurements to a shelter]. These neighboring values were identified by an *n*-by-*n* binary, geographic, connectivity/weights matrix, C; in ArcGIS. In our model if two sampled, time series, homeless–related geolocation, aggregation stratified, observational predictors were neighbors, then cij = 1, and if not, then cij = 0. In ArcGIS two areal units are deemed neighbors if they share a common non-zero length boundary). From our extensive research in the literature we hypothesized that a positive spatial autocorrelated, socioeconomic stratified, homeless, forecast, model output would signify that geographically nearby sampled frequency values of a stratified, behavioral variable (e.g., opiate usage in a teenage pregnant mother) would tend to be similarly aggregated on a map in a specific, dereferenceable, centroid geolocation (e.g., an low, commercial urban parkland).We assumed that in Tampa-Hillsborough County, homeless-related, socioeconomic, explanatory, cluster-oriented, frequency sampled, diagnostic variables may tend to be moderately positively spatially autocorrelated because of the way phenomena are geographically organized in the County. Demographic and socio-economic characteristics like population density and house prices are good examples of variables exhibiting positive spatial autocorrelation in Tampa. For example, neighborhoods in the city tend to be socioeconomically systemized, in such a fashion that clusters of households with similar preferences (e.g., dereferenceable, real estate homestead values >$250,000) are in separate sectors. Homeless populations in Tampa may tend to organize themselves in a way that concentrates similar attributes on a map—creating positive spatial autocorrelation among many socioeconomic explanatory predictor variables—with no government policies nor activities, such as city planning and zoning, thus reinforcing such patterns.

To parsimoniously construct the prognosticative, homeless, frequency-oriented, risk model we first calculated the georeferenced Euclidean distance separating the sampled socioeconomic stratified predictors in geographic space, in ArcGIS, and then squared the difference between their respective feature attribute prognosticative values. Next, distances were grouped into frequency ranges having multiple paired differences, and then group averaged. Subsequently, the distances and the squared attribute differences in the frequency-sampled, homeless, socioeconomic stratified, time series prognosticators were autocorrelated. Our assumption was that semi variance in an empirical stratified dataset of georeferenced, homeless, stratified, time series, dependent, frequency values may equal these squared attribute differences under certain conditions (e.g., if the product is divisible by 2 in a largesampled dataset). Semi variance is a representation in statistics of the analysis of data that fall below the mean value of a set of data. We also assumed that formulating the semivariance in a frequency-oriented, georeferenced, empirical dataset of socioeconomic stratified, temporal sampled, homeless, parameterizable regression estimators may require summing the average of the squared deviations of the sampled variables by their covariate coefficient, discrete, integer values that fell below the mean.

Finally, we assumed that a frequency-oriented, homelessness stratified, discrete, time series, indexed graph whose vertical axis was quantifiable based on an average semivariance, and whose horizontal axis was the averaged Euclidean distance measurements of sampled coordinates may be plotted in space and the distance separating locational predictors [e.g., (*i* and *j*) may be optimally deduced when δij is a binary 0/1 variable. By considering the forecasted regression returns as uncertain variables, proposed a multi-period mean semivariance portfolio optimization malaria habitat model with real-world constraints, in which operational field costs, cardinality and bounding constraints were considered. The authors provide an equivalent deterministic form of mean–semivariance model under the assumption that the predicted habitat of hyper larval productivity included uncertain variables. After that, a modified imperialist competitive algorithm was developed to solve the corresponding optimization problem. Finally, a numerical example was given to illustrate the effectiveness of the proposed habitat model and the corresponding algorithm which was subsequently field validated ("ground truthed') which revealed a 92 percent specificity rate.

As such, here we assumed we would be able to denote whether or not both sampled *i* and *j* belong to a georeferenced homeless stratified, geospatial, temporal cluster. These types of analyses, we assumed, could identify georeferenceable, socioeconomic explanatory co-factors that correspond to the spatial distribution of homelessness clusters in Tampa-Hillsborough County. In this analyses zip code stratified, homelessness rates were the dependent/response variable. Our present study is an attempt to contribute to the literature the continuing integration of a structural and dynamic, forecast-oriented, vulnerability, spatial autocorrelation frequency model of homelessness in Tampa-Hillsborough County, for beginning to answer the “where to target” for planners of homelessness prevention programs. Further by adding to researchers’ tools for investigating the structural aggregation correlates of homelessness (or the “what to target” question facing county planners) other investigators may also exploit county sampled data for implementing sustainable interventions (e.g., mobile needle exchange programs and maternity care, properly geolocated food camps, such as those close to shelters, etc.). In this article our objectives were to a) define geospatialized, frequency clusters stratified by socioeconomic homeless related explanatory covariates using Moran’s *I* statistics and b) to conduct extensive field verification (e.g., video interviews) on victims in hot and cold spot, georeferenced designated geospatial clusters to precisely determine causation covariates of homelessness in Tampa-Hillsborough County, Florida.

Methods and Materials

Hillsborough County is a county in the U.S. state of Florida. In the 2010 census, the population was 1,229,226 making it the fourth-most populous county in Florida and the most populous county outside the Miami metropolitan area. A 2018 estimate has the population of Hillsborough County at 1,436,888 people, which itself is greater than the populations of 12 states according to their 2018 population estimates. Its county seat and largest city is Tampa. Hillsborough County is part of the Tampa–St. Petersburg–Clearwater Metropolitan Statistical Area. According to the U.S. Census Bureau, the county has a total area of 1,266 square miles (3,280 km^2), of which 1,020 square miles (2,600 km^2) are land and 246 square miles (640 km^2) (19.4%) are covered by water. About 158.27 miles (254.71 km) of shoreline are on Tampa Bay.The county’s unincorporated area is around 888 square miles (2,300 km^2), more than 84% of the total land area. Municipalities account for 163 square miles (420 km^2). The modern boundaries of the county place it midway along the west coast of Florida (Figure 1).

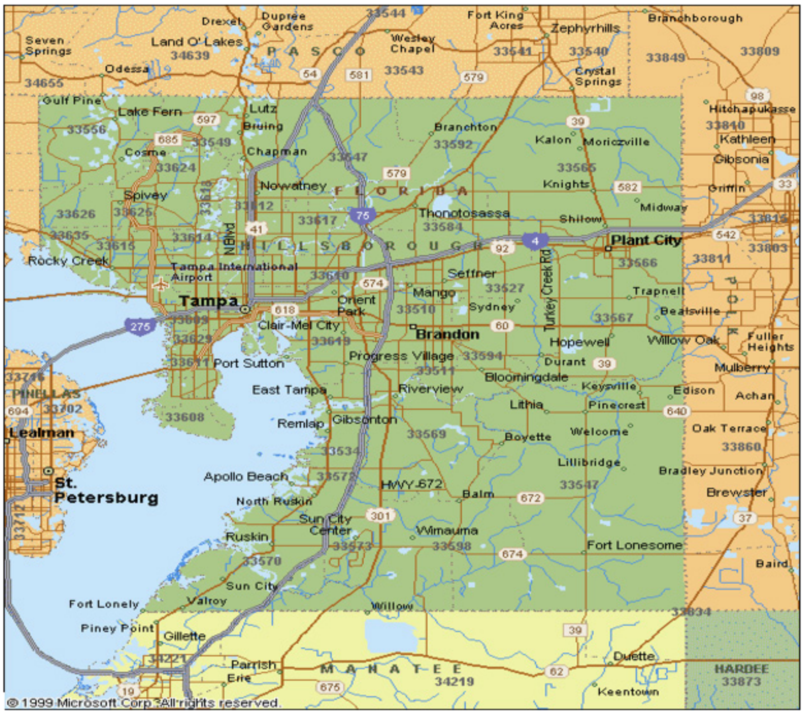

Figure 1: Map of Tampa-Hillsborough County.

Sociodemographic data

According to the 2019 Homeless Count in Hillsborough County, on any given night there are at least 1,650 homeless men, women, and children in Tampa-Hillsborough County. These are people who are sleeping on the streets, behind buildings, in encampments, in cars, emergency shelters and transitional housing. According to the 2019 Homeless Point-in-Time Count, we know the following about who is homeless: 38% are female, 19% are under the age of 18, 10% have served in the U.S. Military, 20% are Hispanic, 18% report experiencing mental illness, and 16% are chronically homeless. Tampa Hillsborough Homeless Initiative [2]. Homelessness happens when a person is unable to afford to pay for a place to live or their current home is unsafe or unstable. In 2018, Hillsborough County, FL had a population of 1.44M people with a median age of 37.1 and a median household income of $58,480. Between 2017 and 2018 the population of Hillsborough County, grew from 1.41M to 1.44M, a 2.01% increase and its median household income grew from $54,731 to $58,480, a 6.85% increase.

Quantitating clusters in geographic space

In order to determine clustering of the GPS-indexed, homelessness populations in Tampa-Hillsborough County, we employed geospatially defined socioeconomic, sampled, frequency variables. For example, we employed the variable (the number of persons who are living at or below the poverty line for the past 12 months) from the ACS Poverty dataset to the homelessness shapefile as a new field in an attribute table in ArcGIS. In this way, the georeferenced, homeless indexed, capture points located in any given census tract in Hillsborough County were also associated with the number of people in that census tract who identified as living at or below the poverty line in the past twelve months. The ACS determined poverty status of the individual by comparing 12 months of income to the poverty threshold. The homeless stratified, GPS dataset was sorted in an ascending pattern based on the time series regressed variables within Hillsborough County.

Explanatory covariates

Potential regression observational socioeconomic predictors of homelessness in this research included 1) Age 2) Race/ethnicity 3) Education level 5) Work history 6) Marriage status 7) Teen motherhood 8) Mental illness 9) Imprisonment and 10) Victim of Domestic violence. Unless otherwise noted, predictors were scored as 1 if present and 0 if not present.

Autocorrelation analyses

The homeless stratified, epidemiologic frequency dataset was stratified into georeferenced groups of population proportions based on their distribution, two standard deviations below and above the median of number of people living below the poverty line. In so doing, the number of individuals above two standard deviations from the median were inferred to be in the low socio-economic georeferenced geolocations in Tampa-Hillsborough County, while individuals at the lower spectrum of two standard deviations below the median were taken to be in the higher socioeconomic zones. Additionally, the assumption for spatial independence was tested for the epidemiologic observations employing the Pearson product moment correlation coefficient [Moran's Index(*I*)]. Moran's *I* was employed as diagnostic tool for quantitating model misspecifications, spatial non-homoscedasticity and outliers in the remotely sensed, parameter estimator, epidemiologic, homelessness, frequency dataset. Homoscedasticity describes a situation in which the error term (that is, the "noise" or random disturbance in the relationship between the independent variables and the dependent variable) is the same across all values of the independent variables. [McCulloch 1985]. Likewise, Moran's *I* was employed to determine if the dependent variables were clustered or randomly distributed within a geographic space in Hillsborough County. We used PROC VARIOGRAM in SAS 9.4 to generate Moran's *I* by computing cross mean of Euclidean inter-site distances between homeless explanatory values that were geographic neighbors. Similarly, the LAGDISTANCE OPTION indicated the neighborhood size in the PROC VARIOGRAM procedure, which was important in the computation of autocorrelation index for quantitating clustering propensities in the sampled homeless clinical and socioeconomic variables. It is of note that lag distance in this research was dependent on the sampled county parameter estimator dataset. Our goal was to create a variogram that invariably provided optimal estimates of spatial dependence for the underlying stochastic process within the dataset. The compute statement allowed averaging of binary spatial weights within the autocorrelation statistical process needed for the construction of Moran's *I* coefficient (an equivalent of regression slope for the Moran's scatter plot). Using the values of LAGDISTANCE=7 and MAXLAGS=10 we constructed a homeless frequency model in Proc Variogram without the Novariogram option in order to compute the empirical semivariogram.

A variogram is often defined as a measure of spatial variability. Our strategy was that by sampling georeferenceable, socioeconomic stratified, capture points close to each other, then this would produce typically similar outcomes compared to sampling for the points separated by larger distances in geographic space. Here the variogram distance measured the degree of dissimilarity $\gamma(h)$ between the sampled, socio economic stratified, homeless data separated by a class of vectors $\boldsymbol{h}$. If $z(x_i)$ and $z(x_j + h)$ were pairs of socioeconomic explanatory samples lying within a given class of distance and direction, then $N(\boldsymbol{h})$ was the number of data pairs within the class. Subsequently, the experimental semi variogram was defined in ArcGIS as average squared difference between the components of the sampled, socioeconomic stratified data pairs in geographic space employing the following equation:(1) γh=12Nh∑i=1Nhzx–xi+h2] This spatial variability measure was a semi variogram. We interpolated between the sample variogram, socioeconomic stratified, explanatory, time series homeless estimators. The variance of the entire data set is referred to as the sill and the distance at which the model semi variogram meets the data set variance is defined as the range.

We specified the CL option in the COMPUTE statement to calculate the 95% confidence limits for the classical semivariance. The Compute Statement described how to use the ALPHA= option to specify a different confidence level in the, homeless, frequency, forecast model. We requested a robust version of the semivariance with the ROBUST option in the COMPUTE statement. Proc Variogram rendered a plot showing both the classical and the robust empirical semi variograms. The Plot option to specify different instances of plots was featured in the empirical semi variogram. In addition, the autocorrelation Moran's *I* statistics under the assumption of randomization using binary weights was generated. The output from the requested autocorrelation analysis included the observed, computed, Geary's c coefficients. The expected value and standard deviation for each sampled homeless, clinical and socioeconomic, stratified, explanatory, covariate the corresponding Z score, and the *p*-value were tabulated in the Pr >j Z j column. The low *p*-values suggested strong autocorrelation for both statistics types. Note that a two-sided p-value was reported, which was based on the probability that the observed, homeless-oriented, frequency coefficients lay farther away from j Z j on either side of the coefficient's expected value—that is, lower than Z or higher than Z. The sign of Z for both Moran's *I* and Geary's c coefficients indicated positive autocorrelation.

The output randomization estimates from the homeless, stratified, autocorrelation, frequency model was then evaluated in a spatial error (SE) model. An autoregressive model was employed whereby a sampled, temporally dependent, socioeconomic stratified, explanatory variable, Y, as a function of nearby sampled homeless-related frequency Y values [i.e., an autoregressive response (AR) or spatial linear (SL) specification] and/or the residuals of Y as a function of nearby Y residuals [i.e., an AR or SE specification].Distance between frequency- sampled, stratified homeless predictors was subsequently defined in terms of an *n*-by-*n* geographic weights matrix, **C**, whose c_{ij} values were 1 if the sampled *i* and *j* were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum, with the row sums given by **C1**, converted this matrix-to-matrix **W.**

The *n*-by-1 vector $x = [x_1 \cdots x_n]^T$ contained measurements of a quantitative, sampled homeless, georeferenced, frequency-oriented explanator for *n* spatial units and *n*-by-*n* spatial weighting matrix **W**. The formulation for the Moran's index of spatial autocorrelation employed in this research was:

$$I(X)=\frac{n\sum_{(2)} w_{ij}(X_i-\overline{X})(X_i-\overline{X})}{\sum_{(2)} w_{ij}\sum_{i=1}^{n}(X_i-\overline{X})^2} \quad \text{where}$$

$$\sum_{(2)}\sum_{i=1}^{n}\sum_{j=1}^{n} \quad \text{with } i \neq j$$

The values w_{ij} were the spatial weights stored in the symmetrical matrix **W** [i.e., ($w_{ij} = w_{ji}$)] that had a null diagonal ($w_{ii} = 0$). The matrix was initially generalized to an asymmetrical matrix **W**. Matrix **W** can be generalized by a non-symmetric matrix W^* by using $W = (W^* + W^{*T})/2$ [see Griffith [1]] Moran's *I* was rewritten in ArcGIS using matrix notation:

$$I(X)=\frac{n}{1'W1}\frac{X^T HHWHHX}{X^T HHX}=\frac{n}{1'W1}\frac{X^T HWHX}{X^T HX} \quad \text{where}$$

$H = (I - 11^T/n)$ was an orthogonal projector verifying that $H = H^2$, (i.e., *H* was independent). Features of matrix **W** for parsimoniously analyzing the time series, sampled, homeless, frequency covariates included that it was a stochastic matrix, which expressed each observed explanatory value y_i as a function of the average of georeferenced zip code geolocation *i*'s and their nearby socioeconomic, stratified, count data covariates. This allowed a single spatial autoregressive parameter, ρ, to have a maximum value of 1.

A SAR model specification was subsequently employed to describe the autoregressive variance uncertainty estimates. A spatial filter (SF) model specification was also used to describe both Gaussian random variables. The resulting SAR model specification took on the following form:

$$Y=\mu(I-\rho)1+\rho WY+\varepsilon$$

(2.1a) where μ was the scalar conditional mean of Y, and ε was an *n*-by-1 error vector whose elements were statistically independent and identically distributed (i.i.d.) normally random variates. The spatial covariance matrix for equation (2.1), using the sampled time series, socioeconomic, stratified, diagnostic, frequency covariates was $E[(\mathbf{Y} - \mu\mathbf{l})'(\mathbf{Y} - \mu\mathbf{l})] = \Sigma = [(\mathbf{I} - \rho\,\mathbf{W}')(\mathbf{I} - \rho\,\mathbf{W})]^{-1}\sigma^2$, where E (●) denoted the calculus of expectations, **I** was the *n*-by-*n* identity matrix denoting the matrix transpose operation, and σ^2 was the error variance.

However, when a mixture of positive and negative spatial autocorrelation was present in model, a more explicit representation of both effects leads to a more accurate interpretation of empirical results [Griffith [1]]. Alternately, the excluded values may be set to zero, although if this is done then the mean and variance must be adjusted. In this research, two different spatial autoregressive, temporal parameters appeared in the spatial, covariance matrix, frequency, homeless model specification, which in the SAR framework became expressable as:

$$\sum=[(I-<\rho>_{diag}W')(I-<\rho>_{diag}W)]^{-1}\sigma^2 \qquad (2.2a)$$

where the diagonal matrix of autoregressive parameters, $<\rho>_{diag}$, contained two sampled parameters: ρ_+ for delineating socioeconomic, covariate pairs in geographic space displaying positive spatial dependency, and ρ. for those pairs displaying negative spatial dependency. For example, by letting $\sigma^2 = 1$ and employing a 2-by-2 regular square tessellation,

$$\Sigma=\left[\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}-\begin{pmatrix}\rho_+&0&0&0\\0&\rho_+&0&0\\0&0&\rho_-&0\\0&0&0&\rho_-\end{pmatrix}\begin{pmatrix}0&\frac{1}{2}&\frac{1}{2}&0\\\frac{1}{2}&0&0&\frac{1}{2}\\\frac{1}{2}&0&0&\frac{1}{2}\\0&\frac{1}{2}&\frac{1}{2}&0\end{pmatrix}\right]^2$$

for the vector $\begin{pmatrix}y_1\\y_2\\y_3\\y_4\end{pmatrix}$, enabled positing a positive relationship between the sampled homeless frequency covariates, y_1 and y_2, a negative relationship between covariates, y_3 and y_4, and no relationship between covariates y_1 and y_3 and between y_2 and y_4. This covariance specification yielded:

$$Y=\mu(I-\rho_+<I_+>_{diag}-\rho_-<I_->_{diag})1$$

$$+(\rho_+<I_+>_{diag}-\rho_-<I_->_{diag})WY+\varepsilon \qquad (2.3a)$$

where $\mathbf{I}_+$ was a binary 0-1 indicator variable which denoted those socio, economic covariates displaying positive spatial dependency, and $\mathbf{I}_-$ was a binary 0-1 indicator variable denoting those sampled covariates displaying negative spatial dependency, using $\mathbf{I}_+ + \mathbf{I}_- = \mathbf{1}$. Expressing the preceding 2-by-2 example in terms of equation (2.3) yielded:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \mu \left[\begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} - \rho_+ \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} - \rho_- \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} +$$

$$\left[\rho_+ \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} + \rho_- \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} \right] \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

If either $\rho_+ = 0$ (and hence $\mathbf{I}_+ = \mathbf{0}$ and $\mathbf{I}_- = \mathbf{I}$) or $\rho_- = 0$ (and hence $\mathbf{I}_- = \mathbf{0}$ and $\mathbf{I}_+ = \mathbf{I}$), then equation (2.3) reduces to equation (2.1) [Griffth [1]]. This indicator variables classification was made in accordance with the quadrants of the corresponding Moran scatterplot generated using the homeless, socioeconomic stratified, time series, frequency covariates sampled in the Tamapa-Hillsborough County study site.

If positive and negative spatial autocorrelation processes counterbalance each other in a mixture, the sum of the two spatial autocorrelation parameters--($\rho_+ + \rho_-$) will be close to 0 [Griffith [1]]. In this research, the Jacobian estimation was implemented by utilizing the differenced indicator temporally dependent, socioeconomic, sampled prognosticative variables ($\mathbf{I}_+ - \gamma\ \mathbf{I}_-$), for estimating ρ_+ and γ with maximum likelihood techniques, and setting $\hat{\rho}- = -\hat{\gamma}\hat{\rho}+$. The Jacobian generalizes the gradient of a scalar valued function of multiple variables which itself generalizes the derivative of a scalar-valued function of a scalar [Cressie 1993]. A more complex specification was then posited by generalizing these binary indicator variables. We employed F: $R^n \rightarrow R^m$ as a function from Euclidean n-space to Euclidean m-space which was rendered employing the Euclidean distance measurements between sampled socioeconomic covariates. Such a function was given by m covariate (i.e., component functions), $y_1(x_1, xn)$, $y_m(x_1, xn)$. The partial derivatives of all these functions were organized in an m-by-n matrix, the Jacobian matrix J of F, which was as follows:

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

This matrix was denoted by $J_F(x_1,..., x_n)$ and $\frac{\partial(y_1,...,y_m)}{\partial(x_1,...,x_n)}$. The i th row (i = 1,..., m) of this matrix was the gradient of the i^{th} component function y_i:(∇y_i). In this analyses $\mathbf{p}$ was a sampled socioeconomic, frequency sampled covariate in R^n and F (i.e., domestic violence in pregnant women count) was differentiable at $\mathbf{p}$; its derivative was given by $J_F(p)$. The model described by $J_F(p)$) was the best linear approximation of F near the georeferenced county capture point $\mathbf{p}$, in the sense that.

$$F(X) = F(p) + J_F(p)(X - p\_ + o(\|X - p\|) \quad (2.4).$$

The spatial structuring of the homeless frequency model was achieved by constructing a linear combination of a subset of the eigenvectors rendered from a modified geographic weights matrix employing (**I** - **11**'/n) **C** (**I** - **11**'/n) that appeared in the numerator of the MC. Spatial autocorrelation can be indexed with a MC, a product moment correlation coefficient [Griffith [1]]. A subset of orthogonal synthetic eigenvectors was then selected with a stepwise regression procedure. Because (**I** - **11**'/n) **C** (**I** - **11**'/n) = **E Λ E'**, where **E** is an n-by-n matrix of eigenvectors and Λ is an n-by-n diagonal matrix of the corresponding eigenvalues, the resulting frequency, homeless, time series model specification was given by: $Y = \mu 1 + E_k \beta + \varepsilon$ (2.5) where μ the scalar mean of Y, Ek was an n-by-k matrix containing the subset of k <<n eigenvectors selected with a stepwise regression technique, and β was a k-by-1 vector of regression coefficients.

A number of the eigenvectors were extracted from (**I** - **11**'/n) **C** (**I** - **11**'/n), which were affiliated with geographic patterns of the temporally dependent socioeconomic, frequency covariates, sampled in the Tampa-Hillsborough County study site, portraying a negligible degree of spatial autocorrelation. Consequently, only k of the n eigenvectors was of interest for generating a candidate set for a stepwise regression procedure. Candidate eigenvector represents a level of spatial autocorrelation which can account for the redundant information in map patterns [Griffith [1]].

The preceding eigenvector properties resulted in û = ŷ and $\hat{\beta} = \mathbf{E}_k'\mathbf{Y}$ for equation (2.3). Expressing equation (2.3) in terms of the preceding 2-by-2 example yielded

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \mu \left[\begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} - \rho_+ \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} - \rho_- \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} +$$

$$\left[\rho_+ \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix} + \rho_- \begin{pmatrix} 0&0&0&0 \\ 0&0&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{pmatrix} \right] \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

Of note is that because the 2-by-2 square tessellation rendered a repeated eigen value. To identify spatially stratified georeferenceable, geospatial, homeless clusters, Thiessen polygon surface partitioning were generated in ArcGIS to construct geographic neighbor matrices, which also were used in the spatial autocorrelation analysis. Entries in matrix were 1, if two sampled, georeferenced, socioeconomic covariates shared a common Thiessen polygon boundary and 0, otherwise. Next, the linkage structure for each surface was edited to remove unlikely geographic neighbors to identify pairs of sampled, homeless, time series, dependent, parameter estimators sharing a common Thiessen polygon boundary. Attention was restricted to those map patterns associated with at least a minimum level of spatial autocorrelation, which, for implementation purposes, was defined by $|MC_j/MC_{max}| > 0.25$, where MC_j denoted the *j*th value and MC_{max}, the maximum value of MC. This threshold value allowed two candidate sets of eigenvectors to be considered for substantial positive and substantial negative spatial autocorrelation respectively. These statistics indicated that the detected negative spatial autocorrelated, socioeconomic stratified, homeless, frequency clusters may be considered to be statistically significant, based upon a randomization perspective. Of note, is that the ratio of the predicted error sum of squares (PRESS) statistic to the sum of squared errors from the MC scatterplot trend line was 1.29 which was well within two standard deviations of the average standard prediction error value (roughly 1.12) for a sampled, explanatory, socioeconomic, stratified, frequency covariate in the Tampa-Hillsborough study site. Because count data was being analyzed, a Poisson spatial filter model specification was employed in this research Detected overdispersion (i.e., extra-Poisson variation) results in its mean being specified as gamma distributed.The model specification was written as follows: where μ_i was the expected mean count for a georeferenced zip code geolocation *i*, μ was an *n*-by-1 vector of expected socioeconomic estimator counts, LN denoted the natural logarithm (i.e., the GLM link function), *α* was an intercept term, and *η* was the negative binomial dispersion parameter. This log-linear equation had no error term; rather, estimation was executed assuming a negative binomial random variable.

The upper and lower bounds for a spatial matrix generated using Morans indices (*I*) was hence given by $\lambda_{max}(n/1^TW1)$ and $\lambda_{min}(n/1^TW1)$ where λ_{max} and λ_{min} which were the extreme eigenvalues of $\Omega = HWH$. Furthermore, the eigenvectors of Ω were vectors with unit norm maximizing Moran's *I* in the frequency, forecast, county homeless, risk model. The eigenvalues of this matrix were equal to Moran's *I* coefficients of spatial autocorrelation post-multiplied by a constant. Eigenvectors associated with high positive (or negative) eigenvalues have high positive (or negative) autocorrelation Griffith [1]. In this investigation orthogonal frequency eigenvectors associated with eigenvalues with extremely small absolute values corresponded to low spatial autocorrelation which were not suitable for defining spatial structures in the sampled, socioeconomic, homeless, stratified covariates.

The diagonalization of the spatial weighting matrix generated from the homeless sampled , zip code, indexed, frequency, socioeconomic stratified, covariate coefficients consisted of finding the normalized vectors u_i, stored as columns in the matrix $U = [u_1 \cdots u_n]$, satisfying the expression $\Omega = HWH = U \wedge U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$ where $\Lambda = diag\ (\lambda_1 \cdots \lambda_n)$, $u_i^T u_i = \|u_i\|^2 = 1$ and $u_i^T u_j = 0$ for $i \neq j$.

Note that double centering of Ω implied that the orthogonal eigenvectors u_i generated from the sampled socioeconomic stratified covariates were centered and at least one eigenvalue was equal to zero. Introducing these eigenvectors in the original formulation of Moran's *I* led to:

$$I(X) = \frac{n}{1^T W1} \frac{X^T HWHX}{X^T HX} = \frac{n}{1^T W1} \frac{X^T U \wedge U^T X}{X^T HX} = \frac{n}{1^T W1} \frac{\sum_{i=1}^{n} \lambda_i X^T u_i u_i^T X}{X^T HX} \quad (2.6)$$

Considering the centered vector $z = Hx$ and employing the properties of idempotence of *H*, equation (2.6) was equivalent to:

$$I(X) = \frac{n}{1^T W1} \frac{\sum_{i=1}^{n} \lambda_i z^T u_i u_i^T z}{z^T z} = \frac{n}{1^T W1} \frac{\sum_{i=1}^{n} \lambda_i \left\|u_i^T z\right\|^2}{\|z\|^2} \quad (2.7)$$

As the eigenvectors u_i and the vector *z* were centered, equation (2.7) was rewritten:

$$I(x) = \frac{n}{1^T W1} \frac{\sum_{i=1}^{n} \lambda_i cor^2(u_i, z)\, var(z) n}{var(z) n} = \frac{n}{1^T W1} \sum_{i=1}^{n} \lambda_i cor^2(u_i, z) \quad (2.8)$$

In this research, *r* was the number of null eigenvalues of Ω ($r \geq 1$). These eigenvalues and corresponding eigenvectors were removed from Λ and *U* respectively. Equation 2.8 was then strictly equivalent to: $I(X) = \frac{n}{1^T W1} \sum_{i=1}^{n-\gamma} I(u_i) cor^2(u_i, z)$ (2.9).Moreover, it was demonstrated that Moran's index for a given eigen vector ui was equal to I(ui)=(n/1T W1)λ i so the equation was rewritten: $I(X) = \frac{n}{1^T W1} \sum_{i=1}^{n-\gamma} I(u_i) cor^2(u_i, z)$ The term $cor^2(u_i, z)$ represented the part of the variance of *z* that was explained by u_i in the homeless county model $z = \beta_i u_i + e_i$. This quantity was equal to $\beta_i^2 / n\ var(z)$. By definition, the eigenvectors u_i were orthogonal, and therefore, regression coefficients of the linear models $z = \beta_i u_i + e_i$ were those of the multiple regression model $z = U\beta + \varepsilon = \beta_i u_i + \cdots + \beta_{n-r} u_{n-r} + \varepsilon$.

The maximum value of *I* was obtained by all of the variation of *z*, as explained by the eigenvector u_1, which corresponded to the highest eigenvalue λ_1 in the spatial autocorrelation error matrix. In this research, $cor^2(u_i, z) = 1$ (and $cor^2(u_i, z) = 0$ for $i \neq 1$) and the maximum value of *I*, was deduced for Equation (2.9), which was equal to $I_{max} = \lambda_1(n/1^TW1)$. The minimum value of *I* in the error matrix was obtained as all the variation of *z* was explained by the eigenvector u_{n-r} corresponding to the lowest eigenvalue λ_{n-r} generated in the homeless frequency model. This minimum value was equal to $I_{min} = \lambda_{n-r}(n/1^TW1)$. If the sampled socioeconomic stratified predictor variable was not spatialized, the part of the variance explained by each eigenvector was equal, on average, to $cor^2(u_i, z) = 1/n\text{-}1$. Because the socioeconomic variables in *z* were randomly permuted, it was assumed that we would obtain this result. In this research the set of *n*! random permutations, revealed that $E_R(I) = \frac{n}{1^T W1(n-1)} \sum_{i=1}^{n} \lambda_i = \frac{n}{1^T W1(n-1)} trace(\Omega)$. It was easily demonstrated that $trace(\Omega) = -\frac{1^T W1}{n}$ and it followed that $E_R(I) = -\frac{1}{n-1}$

Results

We conducted a vigorous spatial autocorrelation vulnerability analyses in ArcGIS employing multiple dereferenceable, socioeconomic stratified homeless, county observational, explanatory predictors. The Spatial Autocorrelation tool returned five values: the Moran's *I* Index, Expected Index, Variance, z-score, and *p*-value in ArcGIS. These values were written as messages at the bottom of the geoprocessing pane during tool execution and passed as a derived output values for potential use in models or scripts. We accessed the data by hovering over the progress bar, clicking on the pop-out button, in the Geoprocessing pane. We evaluated all the autocorrelation data which was accessed as an HTML report file with a graphical summary of results. The path to the report was included with the messages summarizing the tool execution parameters. Clicking on that path opened the report file. Moran's, *I* evaluated whether the county homelessness patterns expressed was clustered, dispersed, or random based on the stratified socioeconomic predictors. When the Z score indicates statistical significance, a Moran's *I* value near +1.0 indicates clustering while a value near – 1.0 indicates dispersion (www.esri.com). The Global Moran's *I* function also calculated a Z score value that indicated whether or not we could reject the null hypothesis. In this case, the null hypothesis stated "there is no spatial clustering in the homeless sampled socioeconomic data of Tampa-Hillsborough County. In this tool, the Z Score was based on randomization null hypothesis computation. A Z-score is a numerical measurement used in statistics of a value›s relationship to the mean (average) of a group of values, measured in terms of standard deviations from the mean. If a Z-score is 0, it indicated that the homeless related data point's score was identical to the mean score. To determine if the Z score is statistically significant, we compared it to the range of homeless county sampled values for a particular confidence level. For example, at a significance level of 0.05, a Z score would have to be less than – 1.96 or greater than 1.96 to be statistically significant in the Hillsborough county model. The Moran's *I* value, and associated Z score were written to the Command window and passed as derived outputs. The input field we selected only contained positive numeric values. Negative weights were converted to zero for the calculations. Thereafter multiple spatial autocorrelation cluster maps were generated (Figure 2, Figure 3, Figure 4, Figure 4A,4B).

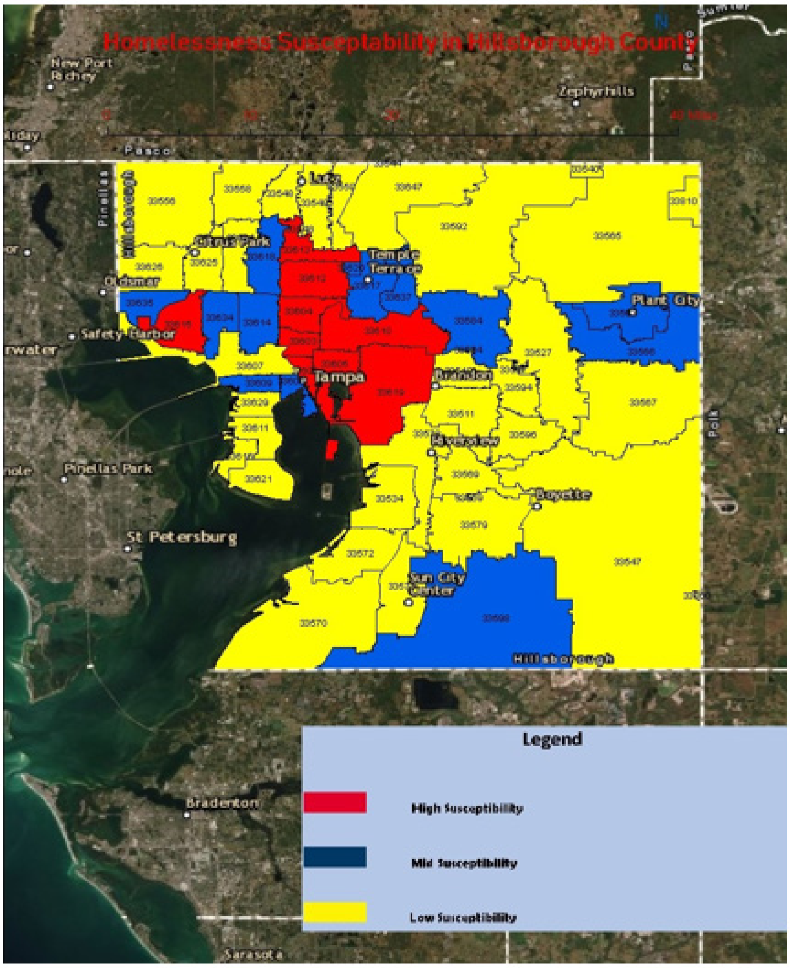

Figure 2A: Final Autocorrelation Homeless Rick Map.

Figure 2B: Ground Trothing highly positive autocorrelated homeless geo referenced cluster.

Figure 3: Conducting interview with a mentally challenged homeless person in a negative autocorrelation cluster during field verification of the socioeconomic cluster model.

Figure 4: Living quarters of a homeless person found a slightly positive autocorrelation cluster.

Figure 4A: Homeless women in a negative autocorrelation cluster.

Figure 4B: A homeless mau in a downtown Tampa positively auto correlated cluster.

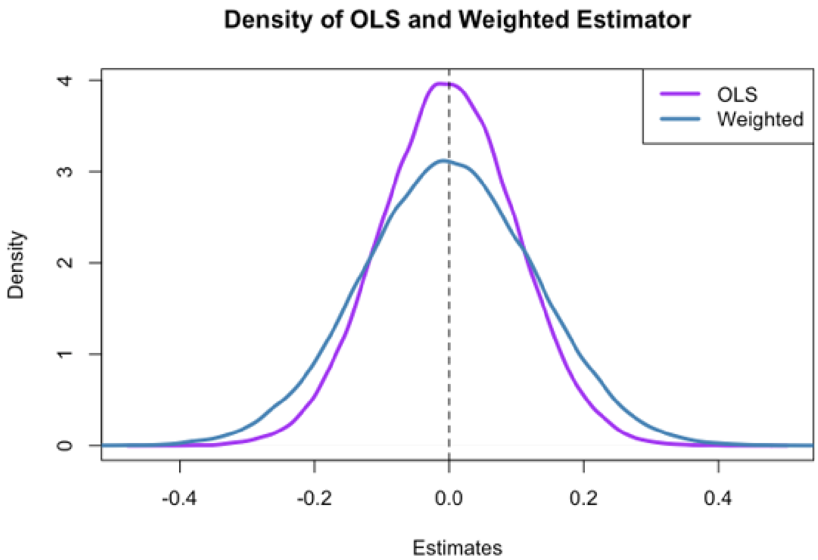

Figure 5: Density of OLS and weighted estimator.

Discussion

We generated multiple spatial auto correlation indexed, homeless stratified frequency clusters for Tampa-Hillsborough County using on-line census and socio-demographic explanatory variables. The analyses computed the mean and variance for the homeless attributes being evaluated. Then, for each feature value, the analyses subtracted the mean, creating a deviation from the mean. Deviation values for all neighboring features (i.e., attribute features within the specified distance band, for example) were multiplied together to create a cross-product. We noted that the numerator for the Global Moran's *I* statistic includes these summed cross-products. Moran's *I* statistic then quantitated the propensities for homeless popualtions to cluster in specific georeferenced socio-economic zones of the county.

We quantitated some degree of spatial autocorrelation amongst the spatially distributed, univariate, homeless, frequency socioeconomic observations. This autocorrelation we assumed originated from (a) missing exogenous factors that exhibited distinctive spatial patterns in homeless socioeconomic time series datasets and thus geographically tied the residuals together, or (b) underlying spatial processes that emerged from spatial exchange mechanisms among the regressors; and/or, (c) an inappropriate spatial aggregation of the underlying observational units. The presence of spatial autocorrelation violates the ordinarily stated assumption of stochastic independence among socioeconomic stratified observations, on which statistical inference from most classical statistical frequency models is based in the literature. Thus, ignoring spatial autocorrelation in these paradigms can lead to biased standard errors and/or biased parameter estimates, as well as artificially inflated degrees of freedom. This would result in skewed data outputs due to spatial heteroskedasctcity.

Common practice, when dealing with spatially distributed observations, is to use either maximum likelihood or Bayesian estimation. In this study, we investigated the necessary condition for consistency of the maximum likelihood estimator (MLE) of a spatial homeless, socioeconomic stratified, frequency model with a spatial moving average process in the disturbance term. We show that the MLE of an eigen decomposed dataset of frequency-oriented, spatially autoregressive, socioeconomic stratified, homeless parameters are generally inconsistent when heteroskedasticity is not considered in the estimation. We also show that the MLE of socioeconomic dependent parameters of exogenous, homeless, time series, frequency sampled variables is inconsistent and determine its asymptotic bias. Asymptotically unbiased estimators are operators whose bias goes to 0 as the explanators (e.g., time series frequency stratified sample size of socioeconomic homeless variables in regression space goes to infinity. According to Jacob if $\hat{\theta}_n$ is an autocorrelated estimator of θ using a sample of size *n*, then this estimator is asymptotically unbiased if $\lim_{n\to\infty} E[\widehat{\theta}_n|\theta] = \theta$

We provide simulation results to evaluate the performance of the MLE. The simulation results indicated that the MLE imposed a substantial amount of bias on both autoregressive and moving average temporally dependent, homeless, stratified, socioeconomic parameter estimators in the forecast model. These parametric estimators explicitly specified the distributional characteristics of the underlying models. In contrast, nonparametric methods are distribution free without sacrificing too much information in a sample. However, these models may become quite computer intensive. See Hollander and Wolfe for more details on nonparametric statistical methods. More specifically in the spatial domain, the nonparametric eigenvector filtering procedure does not require restrictive and perhaps unjustified distributional assumptions when constructing predictive, vulnerability-oriented, socioeconomic stratified, homeless frequency, county-level, cluster models.

The eigenvector spatial filtering procedure is founded on the standard ordinary least squares (OLS) estimator and is, apart from the assumptions of independence and constant variance of the disturbances, distribution free owing to the Gauss Markov theorem (see Appendix 1). The spatial filtering estimator is fairly robust to model specification errors for optimizing homeless frequency predictive risk modelling county sampled, temporal socioeconomic stratified variables compared with a spatial MLE. The interpretation of its results is straightforward as the different components of a spatial process can be extracted and visualized. If a homeless data analyst or epidemiologist needs to preserve some structural properties of spatial models, then spatial filtering can be implemented as a semiparametric method.

In this paper we concentrated on standard linear regression models y=Xb+e, where y is an (n>1) vector of the endogenous variable for the n georeferenced socioeconomic observations, X is an (n=k) matrix of k exogenous variables, including an (n>1) unity vector 1, bis the (k>1) vector of regression parameters, and is an vector of random disturbances. We assumed that spatial autocorrelation among regression disturbances was induced by exogenous spatially autocorrelated socioeconomic co-factors, which were not incorporated into the homeless, frequency, county model. This led to a model misspecification by shifting parts of the relevant information from the mean response X (or first-order component) into an (*nxn*) covariance structure of the disturbances [or second-order component cov.].

Alternatively, we may allow an underlying spatial process in a homeless frequency, time series, epidemiological, diagnostic model which may be induced by spatial autocorrelation. Furthermore, an observed spatial pattern in the response variable in such a paradigm may be decomposable into, preferably three, statistically independent components: (a) a systematic spatial trend component that is specified by a parsimonious set of exogenous variables with a substantive meaning for the problem under investigation; (b) a stochastic signal that reflects either an underlying spatial process and/or a set of missing exogenous factors with an inherent spatial pattern; and (c) the independent white-noise disturbances. In discrete time, white noise in a homeless forecast, vulnerability, county model is a discrete signal whose samples (i.e., socioeconomic regressors) are regarded as a sequence of serially uncorrelated random variables with zero mean and finite variance . Depending on the context, an epidemiologist or researcher may also require that the samples be independent and have identical probability distribution (in other words independent and identically distributed random socioeconomic variables would be the simplest representation of white noise) in a homeless model In particular, if each sample has a normal distribution with zero mean, the signal would be classified as additive white Gaussian noise.

Hence, a random homeless vector (that is, a partially indeterminate process that produces vectors of real discrete integer values) is said to be a white noise vector or white random vector if its components each have a probability distribution with zero mean and finite variance, and are statistically independent: that is, their joint probability distribution must be the product of the distributions of the individual components. A necessary (but, in general, not sufficient) condition for statistical independence of two variables is that they be statistically uncorrelated; that is, their covariance is zero. Therefore, the covariance matrix R of the components of a white noise vector w with n elements in a time conscious, homeless forecast, epidemiological model must be an n by n diagonal matrix, where each diagonal element R_{ii} would be the variance of component w_i; and the correlation matrix would be be the n by n identity matrix. If, in addition to being independent, every sampled homeless stratified socioeconomic explanatory variable in w also has a normal distribution with zero mean and the same variance {\displaystyle \sigma ^{2}}, w would be be a Gaussian white noise vector. In such cases the joint distribution of w in the paradigm would be a multivariate normal distribution; the independence between the variables would imply that the distribution has spherical symmetry in n-dimensional space. Therefore, any orthogonal transformation of the vector will result in a Gaussian white random vector in the model outcome.

Often the weaker condition "statistically uncorrelated" is used in the definition of white noise, instead of "statistically independent". However, some of the commonly expected properties of white noise may not hold for this weaker version in a homeless predictive county risk model. Under this assumption, the stricter version can be referred to explicitly as independent white noise vector. Other authors use strongly white and weakly white instead. An example of a random vector that is "Gaussian white noise" in the weak but not in the strong sense is $x=[x_1,x_2]$ where x_1 is a normal random variable with zero mean, and x_2 is equal to $+x_1$ or to $-x_1$, with equal probability. These two variables are uncorrelated and individually normally distributed, but they are not jointly normally distributed and are not independent. If x is rotated by 45 degrees, in a homeless model its two components will still be uncorrelated, but their distribution will no longer be normal. In some situations an experimenter may relax the definition by allowin geach component of a white random vector w to have non-zero expected value in the epidemiological, prognosticative, county, homeless model. {\displaystyle \mu }The underlying rationale for the eigenvector spatial filtering approach for homeless, predictive, risk modelling socioeconomic explanatory sampled variables is eigenvectors that are extracted from a transformed spatial link matrix exhibit distinctive spatial patterns with associated spatial autocorrelation levels. Furthermore, these eigenvectors are mutually orthogonal and uncorrelated. A linear combination of a small subset of these eigenvectors are capable of: (a) capturing the hidden spatial pattern of a stochastic component in a homeless county frequency model, and (b) filtering this pattern from the covariance matrix of the sampled socioeconomic, stratified, frequency, time series, georeferenced observations. Thus, this subset of eigenvectors is a proxy either for those spatially autocorrelated exogenous socioeconomic co- factors that have not been incorporated into a homeless, temporally dependent, frequency model, or for an underlying spatial process that ties the sampled observations together in geographic space. Furthermore, incorporation of all relevant eigenvectors into a homeless frequency socioeconomic stratified time series model would leave the remaining residual component spatially uncorrelated. Consequently, standard statistical modeling and estimation techniques as well as interpretations can be optimally employable for constructing county-level, spatially filtered homeless, predictive risk models.

The key theoretical and practical issues of the eigenvector filtering approach for time series homeless forecast vulnerability, cluster modelling is: (a) which eigenvectors constitute the potential candidates for specific regression models, and (b) which selection strategy leads to spatially uncorrelated regression residuals and a parsimonious set of eigenvectors. Here we proposed alternative collections of eigenvectors that allowed us to perform semiparametric spatial filtering on an empirical dataset of time series sampled dereferenceable, socioeconomic stratified variables in geographic space. We investigated more closely different selection strategies that can be employed to derive the most parsimonious subset of eigenvectors. Initially we reviewed parametric as well as nonparametric spatial filtering methods for optimally quantitating the temporal sampled, georeferenced, homeless frequency, socioeconomic, stratified, parameter estimators. Next, several spatial autoregressive regression zip code models were constructed. We then spatially linked the semiparametric spatial filtering methods and the autoregressive spatial regression models. Finally, stepwise selection strategies for the set of relevant eigenvectors, rendered from the eigenfunction eigen decomposition orthogonal framework was conducted employing their underlying geographic algorithms which allowed cartographically delineating the positive and negative autocorrelated stratified georeferenced clusters.

Thereafter we conducted video interviews to determine causation in positive and negative autocorrelated clusters. Violence, especially for female victims was common in the positive autocorrelated cluster. Mental health was the primary covariate in the negatively autocorrelated cluster. In the positively autocorrelated georeferenced cluster drug usage and transaction was a primary covariate. The top causes of homelessness among unaccompanied individuals in highly positive autocorrelation were (1) unemployment, and 2) drug usage/ transaction. In slightly positive autocorrelated cluster causation was identified as 1) previous incarceration**,** 2) medical care/ food shortage and 3) domestic violence especially for female victims. Mental Health was the primary covariate in the negatively autocorrelated clusters**.**

Credible estimates of the prevalence of alcohol and drug abuse suggest that alcohol abuse affects 30% to 40% and drug abuse 10% to 15% of homeless persons. A review of policies that address substance abuse among the homeless in Tampa-Hillsborough County may reveal interventions alternate between control and rehabilitation. However, the unique needs of a changing homeless population in the county may require an integration of alcoholism and drug abuse recovery services with programs for women, adolescents, and the mentally ill. Alcohol- and drug-free housing may be essential to support and maintain recovery in Tampa-Hillsborough County. Unemployment led the list of causes of homelessness among individuals, followed by lack of affordable housing and lack of needed services, and substance abuse and lack of needed services. The increasing numbers of people leaving carceral institutions face an increased risk for homelessness in the county.

In the negatively autocorrelated clusters, mental illness was determined to be a predominant factor. Since these individuals tended to be socially isolated, they were no evidence of people spatially clustering in Tampa-Hillsborough County. People who are homeless may be vulnerable to myriad health and social problems, which may be exacerbated by the presence of mental illness. People with severe mental illness may become homeless as a direct result of the symptoms of their illness. As a consequence, these people may not utilize social and economic networks, or both in the county. The experience of homelessness may precipitate and exacerbate symptoms of mental illness, whether alone or in the context of substance misuse. The prevalence of serious mental illness is higher in homeless people compared with those who are housed, and there are higher rates of personality disorder, self-harm and attempted suicide.

Homelessness disproportionately affects women and children in Tampa-Hillsborough County. Homeless women are at higher risk of having chronic illnesses, infectious diseases, substance abuse problems, mental illness, and being a victim of sexual or domestic violence more than women who are not homeless (National Coalition for the Homeless [3]). They are also less likely to have insurance, social support, income, or access to preventive health services. Persons who experience homelessness may be less likely to engage in the health care system due to challenging relationships with health care providers, inconvenience, cost, and a perceived lack of compassion and discrimination on the part of the providers. Senior members of the health care team are particularly responsible for educating house staff and students on how to appropriately care for this vulnerable group of women.

During our field verification exercises of our homeless predictive risk mapping variables, we noted that pregnant homeless women were typically younger than non-pregnant homeless women. During the interview process we noted that these victims had a history of family disruption. We were able to surmise that the homeless pregnant teenagers in Tampa-Hillsborough County are commonly the product of domestic instability and/or poverty. In a survey of women at emergency departments and primary care clinics, pregnant women who were homeless had higher rates of cigarette smoking, lower rates of employment, and lower achieved educational levels as compared with consistently housed counterparts Crawford [4]. In a study that reviewed trends of homeless women in the United States from 2000 to 2007, homelessness was associated with Black and Hispanic races, being unmarried, uninsured, and receiving government aid. The prevalence of homelessness among women in this study was 4 percent, which is approximately 1 in every 26 women of reproductive age in the United States. The explanatory frameworks for women's homelessness and its gendered nature, require more invasive research in order to identify major trends in Tampa-Hillsborough County which may have to be addressed differently rather than men homelessness. Research on women's homelessness in the county may provide an interesting and particularly useful approach to the 'construction of homelessness', its practices, socially perceived images and discourses, while offering important insights on policies and practices. Perceptions of homeless women may increase understanding of the complex interactions between power structures and individual agencies in Tampa- Hillsborough County.

We were able to determine other co-factors associated with homelessness in Hillsborough County. For example, we were able to determine that in the county there was a tendency for people to lose jobs and then housing. Homeless women ran away to the street to escape domestic violence. Many people in the county experienced significant trauma and simply could not cope with life. Others struggles with mental illness, depression, or post-traumatic stress of affordable housing in homeless. Once homeless, the lack of housing, access to healthcare, and supportive services, then seem to act as other barriers that keep individuals from moving back into housing.

Concerns with automatic variable selection procedures are well known Chatfield [5] and generally applied in this research effort. Of particular concern is that all candidate eigenvectors coefficients in the homeless frequency analyses had nonzero prior probability for inclusion in the final model but all that failed to meet the threshold for inclusion which was then given by the quantized coefficients of βEi = var(βEi) = 0 which self-evidently eliminated some amount of uncertainty that remained after regressing the socioeconomic data. A proper accounting for uncertainty in the spatial process is important not just in its own right, for the purposes of prediction and possibly cluster detection, but also is necessary for an accurate estimation of the marginal effects of covariates of interest in county homeless frequency model constructed employing empirical socioeconomic stratified dataset of time series, sampled covariates. A primary challenge for ESF estimation is the proper calculation of uncertainty across the high-dimensional parameter space. Jacob explicitly states that within the frequentist paradigm the variance of a multiple regression coefficient estimates βj var (ˆ βj) = σ^2 ∑n i=1(Xji − ¯ Xj)2(1 1−R^2 j) (3) where σ^2 is the residual variance and R^2

j is the R^2 statistic from a regression of xj on the remaining covariates in the design matrix X. Hence the precision of βj increases with the variance of xj and decreases with σ^2 and R^2 j. If μ =Xβx with βx a vector of unknown coefficients in a homeless, georeferenced frequency-oriented socioeconomic stratified geospatial cluster, then any non-zero pairwise correlation between sampled covariates and their eigenvectors necessarily may induce additional uncertainty into the estimate of βx in the model renderings. At the same time as variable selection procedures drop, some homeless frequency, estimator eigenvectors with non-zero correlations derived from the model, may require unpenalized additions. Here these eigenvectors correlated with the outcome variable but decreased the residual variance of the regression. Automatic variable selection procedures in the homeless model construction phase thus worked systematically to minimize the standard error of regression coefficients but without accounting for all sources of uncertainty. The challenge is obviously that a high-variance frequency, homeless, socioeconomic stratified county forecast, vulnerability-oriented, epidemiological model with a predetermined number of *n* estimated eigenvector coefficients may fail to meaningfully delineate the researchers state of knowledge since it ignores a great deal of prior information regarding the degree of complexity in most spatial processes and may over or under-correct for spatial autocorrelation.

In the future, spatial sampling autocorrelation cluster, frequency homeless, socioeconomic stratified risk models may be employed to measure populations that are not straight forward in order to capture homeless groups in Tampa-Hillsborough County that are often underestimated. For example, because of varying definitions of homelessness and transient nature of the homeless populations autocorrelation methods may arrive at estimates of population counts by extrapolating parts of the populations that can be observed and measured. These paradigms can be used to either guide new data collection or estimate population size using existing survey homeless county socioeconomic stratified data.

Time series spatial epidemiological models may be geared towards producing accurate forecasts in outcomes of interest (e.g., causation of geospatial clusters stratified by homelessness socio-economic, explanatory, georeferenced variables) in the short to medium-term based on past trends in Hillsborough County. In order to generate short-time forecasts, these models may depend heavily on the latest observations in the sample. On the other hand, when applied for medium term predictions, they can be adjusted to place more emphasis on longer term homeless trends in Hillsborough County.

Forecasts of urban zones using socio-economic or demographic homeless-related explanatory, temporally dependent georeferenceable variables may be synthesized from the Autoregressive Integrated Moving Average (ARIMA) model, the error-correction model, the Autoregressive Conditional Heteroscedastic (ARCH) model and a Box-Jenkins multi-variate time series analysis. The Arima approach may be the simplest method applied to forecast homelessness, socioeconomic-related trends in Hillsborough County. It may be used to measure links between key predicting factors and future outcomes of homelessness. Branas employed ARIMA techniques to forecast suicide rates conditional on adverse economic conditions while Chamlin applied the Arima methodology to explore temporal relationships between crimes and arrest rates. The Box-Jenkins and Arch approaches, which can be thought of as extensions to the basic ARIMA method, mainly rely on the same principles. They may be also applied to forecasting other homeless, socioeconomic indexed explanatory outcomes such as income, inequality, and poverty. Moreover, the error correction model may arrive at prognostications of welfare outcomes considering their relationships with a set of homeless-related county sampled cluster covariates over time.

Machine learning methods have recently been used to produce projections of welfare outcomes, such as poverty. These methods may identify patterns of connections between explanatory socio-economic co-factors and homeless outcomes in Tampa-Hillsborough County through iterative processes by prioritizing georeferenceable, temporally dictated cluster covariates related to homelessness. High-order interactions between predictive homeless explanatory variables and outcomes of interest that do not need to be specified in advance may be explored. Machine learning methods may be subsequently employed to produce projections of welfare outcomes such as unemployment. Machine learning methods may also map links from predicting homeless georeferenced clusters in Tampa-Hillsborough County to outcomes of interest (geolocations for homeless drug usage intervention). These paradigms may be effectively applied when the goal of the research is an accurate prediction of specific outcomes rather than the estimation of separate effects of single causal factors in the county [6-10].

With the ArcGIS platform, organizations can gain a high-level overview of all operations in real time, and in one place for mapping county homeless populations in Tampa-Hillsborough County. Operations Dashboard for ArcGIS is a ready-to-use application that can configure any outreach homeless–related socioeconomic program so that a researcher or epidemiologist can focus on what matters most. Interactive maps and data sources can be updated automatically as field information changes so researchers can view a current homeless, frequency- stratified cluster location for outreach teams where surveys could be administered. Furthermore, the dashboard may determine what types of issues are becoming apparent, and where data is missing. Shelter staff can access real-time data to know how many people they are serving, while administrators can obtain a current view of shelter capacity across a given region of Tampa-Hillsborough County. Giving executives visualizations of homeless frequency county trends may allow them to be better prepared for public inquiries and budget appropriately.

In future research, the spatial autocorrelation county cluster models should include time of early substance exposure, self-esteem, and other psycho-social variables. Such analysis may illuminate heterogeneity in homeless, time series stratified, socioeconomic data at the county level. Other frequency variables that should be employed when risk modelling homelessness in Hillsborough County may include neighborhood acceptance and systemic stigma/discrimination which may have substantial influence on housing outcomes.

Strengths of an ArcGIS spatial autocorrelation analysis include usage of data from a large heterogeneous multisite, zip code, county sample with low attrition and comprehensive measurement. Despite these strengths, we caution that "trajectory analyses represent statistical approximations rather than identifiable 'types. Causal inferences are tentative. Our spatial modelling approach is aimed at finding longitudinal patterns in data, and the results which may not be entirely consistent with classical county-level homeless analyses. Even though we specified an autocorrelation model in advance, the analysis was very exploratory, and the class compositions should not be considered absolutely precise. Some socioeconomic sampled variables were not comprehensively assessed. Unfortunately, self-reported homeless variables may be prone to numerous biases. Our suggestions for research for Hillsborough County include longer follow-up, vulnerability modeling variable groupings as higher-level dimensions, and examining other outcomes (e.g., recovery time frames). Further description of specific patterns of housing based on more refined socio-economic zones (lower residential urban within 2 km Euclidean distance of a georeferenced drug overdose death) is also warranted.

Correlations between eigenvectors of a spatial connectivity matrix and covariates have been coined 'spatial confounding'. In contrast to the perspective just outlined, Hodges and Reich argue that the variance inflation caused by correlations between spatial random effects and covariates does not reflect any legitimate inferential uncertainty and can also ''mess up'' the fixed effects estimates obtained from a non-spatial linear model (e.g., homeless frequency, socioeconomic stratified, negative binomial regression with a non-homogenous gamma distributed mean). This view motivated to conduct a search for a model in which ''sample size can be discounted without distorting the fixed effect estimate. To this end Hodges and Reich proposes Restricted Spatial Regression (RSR), a spatial filtering method which introduces the eigenvectors of Moran's eigenvectors where $M = I - X(X'X) -1X'$, so that the eigenvectors are restricted to the space orthogonal to X.

Moran's homeless, frequency, time series, indexed, eigenvectors maps are attractive mathematical objects as they are fairly simple to calculate and can be used in most studies of spatially-explicit socioeconomic, homeless data. There is, however, an aspect of eigen-analysis that still requires some investigation for precision homeless risk, mapping: the effect of irregular cluster sampling on modeling performance. A study may be conducted to investigate empirically the behavior of varying time series stratified irregularity schemes, Moran's eigenvector maps generated from sampled, georeferenced frequency datasets of explanatory socioeconomic stratified, georeferenced, geospatial, cluster covariates may reveal aggregation propensities in the diagnostic estimators. By focusing on simulated scenarios in ArcGIS sampling designs may be usable to determine frequency-oriented fluctuating and constant, causation, homeless parameters. Moran's eigenvector homeless frequency county maps may be computed and correctly used with time series socioeconomic stratified data coming from irregularly designed sampling surveys, given some precautions. Homeless georeferenced county sampling sites may be equally spaced but may not cover an entire county study area, however the Moran's eigenvectors can be still computed directly based on the coordinates of the sampling sites without any important loss of information. Whereas, when the phenomenon of interest is resolved employing randomly stratified sampling designs, the homeless, frequency, socioeconomic stratified, temporally sampled, Moran's eigenvector frequency maps should be computed on a reconstructed space of regular sampling sites followed by removal of the missing sites, before analysis. This solution of rebuilding a (regular) sampling space may capture the underlying process causing the clustering tendencies in the sampled socioeconomic, georeferenced, county, frequency covariates hence improving the modeling results and relaxing the impact of the choice of the weighting matrix on the computation of Moran's eigenvector maps.

RSR is designed such that no correlated eigenvectors are allowed to 'steal' from the explanatory power that an OLS regression would apportion exclusively to the covariates. To the extent that correlations between georeferenced, sampled, frequency stratified, socioeconomic, time series, vulnerability county covariates and their eigen decomposed eigenvectors produce a challenge for probable inference – particularly that of simultaneously estimating a nonstationary mean and the effects of covariates with limited information – surely it would be necessity to address this phenomenon in homeless time sensitive, socioeconomic, parameterized, estimator models. One reason to remain skeptical of RSR is its substitution of a deterministic separation procedure designed to purge a source of parameter uncertainty from the model, for established inferential methods. The notion that the spatial trend component of a frequency diagnostic, homeless, socioeconomic, stratified county model should have no impact on non-spatial parameter estimates overlooks the geographic analog to Yule's ''nonsense correlations'' which routinely appear in nonstationary time series data and motivates much foundational work in spatial statistics. This suggests that the a priori restriction of the space spanned by the spatial filter is counterproductive in the sense that it ignores useful, potentially critical information. Lastly, Murakami and Griffith's RE-ESF model, which was an extension of Hughes and Haran's Bayesian estimation procedure, is a highly promising development for ESF since it produces penalized estimates of βE with empirical Bayes techniques rather than resorting to variable selection methods. However, calculating the degrees of freedom for REML, homeless, socioeconomic stratified, cluster frequency, county temporal models may be a challenge. The restricted (or residual, or reduced) maximum likelihood (REML) approach is a particular form of maximum likelihood estimation that does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data, so that nuisance parameters have no effect [11-13].

The success of homeless outreach programs depends on government agencies, nonprofits, and the community working together in Tampa-Hillsborough County to accomplish the same mission. To determine how to best use internal and external resources, directors and executives at all levels need a sound understanding of how their programs are performing. Being able to understand operations in real time allows county organizations to stay accountable while delivering the right resources to the people who need them most-while getting the job done faster.

Conclusion

In conclusion, our spatial autocorrelation frequency models suggest new supportive facilities and shelters for the homeless should be located in areas with a high availability of employment, inexpensive or free medical care and food in Tampa-Hillsborough County. Furthermore, free mobile drug addiction programs, and family domestic violence interventions should be implemented in the county. Homelessness prevention in the county must not only include interventions targeted at individuals, but broader structural reforms directed at addressing the drivers of homelessness. With intelligent mapping and survey tools, Tampa-Hillsborough County homeless-related agencies can do more with field collected and socio-demographic/socioeconomic data. Geographic Information System (GIS) technology and autocorrelation statistics can deliver the power of geography and analysis to help human service organizations in the county collect, manage, visualize, and understand this data in new ways. Even without a physical address, the location of homeless individuals has a key role in identifying patterns and trends in Tampa-Hillsborough County. Then, communities in the county can truly understand homelessness, see where the need is greatest, and determine the best approach for connecting people with critical resources Figure 5.

Conclusion derivable from the Gauss-Markov Theorem

- Both estimators seem to be unbiased: the means of their estimated distributions are zero.
- The estimator using weights that deviate from those implied by OLS is less efficient than the OLS estimator: there is higher dispersion when weights are wi=1±0.8100wi=1±0.8100 instead of wi=1100wi=1100 as required by the OLS solution
- A number of different approaches have been proposed in the literature for modeling the unmeasured spatial autocorrelation, including geostatistical models simultaneous autoregressive models (Kissling and Carl, 2008), and spline-based models. However, by far the most common approach is to use conditional autoregressive (CAR) models, which are a special case of a Gaussian Markov random field (GMRF). These models represent spatial closeness via an $I \times I$ neighborhood or adjacency matrix **W**, where element w_{ir} defines whether areas (i, r) are spatially close. Typically, a binary specification is used, so that $w_{ir} = 1$ if areas (i, r) are spatially close and $w_{ir} = 0$ otherwise. This specification leads to a sparse specification for **W**, which makes the fitting of these models much more efficient than if **W** was a dense matrix. Commonly, border sharing is used to determine **W**, so that $w_{ir} = 1$ if areas (i, r) share a common border, and $w_{ir} = 0$ otherwise. Given this neighborhood matrix, CAR models for a vector of random effects $\boldsymbol{\phi}$ are most often written as a set of univariate full conditional distributions, $f(\phi_i|\boldsymbol{\phi}_{-i})$, where $\boldsymbol{\phi}_{-i} = (\phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_I)$. However, this set of I conditional distributions is equivalent to the following multivariate Gaussian joint distribution ϕ∼N(0,τ2Q(W)−1), where **0** is an $I \times 1$ vector of zeros and **Q**(**W**) is an $I \times I$, potentially singular, precision matrix. The simplest CAR model is the *intrinsic* model (ICAR, Besag et al., 1991), which is given by ϕi|ϕ−i∼N∑r=1Iwirϕr∑r=1Iwir, τ2∑r=1Iwir, and corresponds to the singular precision matrix **Q**(**W**) = diag (**W1**) −**W** in the above joint specification. This model can capture spatial autocorrelation in a time series sampled, empirical sampled dataset of georeferencable, socioeconomic, stratified, homeless parameter estimators because the conditional expectation would be the mean of the random effects in neighboring areas, while the conditional variance would be inversely proportional to the number of neighboring areas. This latter specification makes sense if the homeless data are spatially autocorrelated, because the more neighbors' area i has with similar random effect values, then the more information and hence the less uncertainty there is about the value of ϕ_i. However, GMRF/CAR model corresponds to an improper joint distribution for $\boldsymbol{\phi}$ with a singular precision matrix, and also only allows for strong spatial correlation that can sometimes lead to over smoothing. Therefore the *convolution* or *BYM* model was proposed by Besag .which augments the intrinsic model with a second set of spatially unstructured random effects.

(5)ϕi=ϕi(1)+ϕi(2),ϕi(1)|ϕ−i(1)∼N∑r=1Iwirϕr(1)∑r=1Iwir,τ2∑r=1Iwir,ϕi(2)∼N(0,σ2).

This model represents the random effects $\boldsymbol{\phi}$ with a convolution of spatially autocorrelated and spatially unstructured effects, which are modeled by the intrinsic CAR model and a zero-mean Gaussian shrinkage model, respectively. This is the most commonly used CAR model in the literature and can induce varying levels of spatial autocorrelation by varying the amount of variation in each of the two components. Two alternative model frameworks have been proposed, which each have a single set of random effects but introduce a spatial dependence parameter ρ to allow for varying levels of spatial autocorrelation. The first was proposed by Stern and Cressie and is given by (6) ϕi|ϕ−i∼Nρ∑r=1Iwirϕr∑r=1Iwir, τ2∑r=1Iwir, and corresponds to a joint distribution with precision matrix **Q** (**W**, ρ) = diag(**W1**) − ρ**W**. Here $\rho = 1$ simplifies to the intrinsic model while $\rho = 0$ corresponds to independence, the latter being the case as then the conditional expectation does not depend on the random effects in other areas. One downside of this model is that if $\rho = 0$ then the conditional variance still depends on the number of neighboring areas, even though there is no spatial autocorrelation in the random effects. Therefore an alternative was proposed by Leroux which is given by(7) ϕi|ϕ−i∼Nρ ∑r=1I-wirϕrρ∑r=1Iwir+1−ρ,τ2ρ∑r=1Iwir+1−ρ, and corresponds to a joint distribution with precision matrix **Q**(**W**, ρ) = ρ[diag (**W1**) −**W**] + (1 − ρ)**I**, where **I** is an identity matrix. In this model $\rho = 1$ simplifies to the intrinsic model, while $\rho = 0$ corresponds to independence with mean zero and a constant variance. More recently, Riebler proposed an alternative CAR model to the above that accounts for scaling, and a number of other extensions are discussed in the remainder of this chapter. Inference for this model is typically undertaken in a Bayesian setting.

A summary of the inferential spatial autocorrelation approaches for predictive homeless modelling may be provided by lattice data, whereby values, (y_1, y_2, ..., y_n) of some socioeconomic stratified, time series, explanatory variable Y are recorded for each member of a set of n areal units (a_1, a_2, ..., a_n) which, taken collectively can cover the study region A, i.e., $a_1 \cup a_2 \cup \ldots \cup a_n = A$. The set of data values may subsequently be considered to represent one possible realization of a spatial process operating over A. Lattice data are analyzed by examining characteristics of the association between pairs of data values as some function of their spatial association. (i.e., spatial autocorrelation).' Since the data sites are homeless aggregated areas, there are many different ways in which the spatial association w_{ij} between any sampled socioeconomic data sites a_j may be modeled. By far the most frequently used method is to set $w_{ij}=1$ if a_i, a_j share a common boundary, and $w_{ij}=0$, if they do not. An observed value of I which is larger than its expectation under the assumption of spatial independence indicates positive spatial autocorrelation (i.e., similar values of y, are found in spatial juxtaposition). In contrast, negative spatial autocorrelation occurs when neighboring values of y_i are mutually dissimilar, indicated by an observed value of Moran›s I which is smaller than its expectation. For data sets of $n \geq 50$, I is approximately normally distributed so that the standard score may be used to determine if the empirical pattern displays significant spatial autocorrelation. The measurement of spatial autocorrelation can be extended to higher-order spatial neighbors where the spatially associated areas are (k−1) intervening areas apart. A plot of the values of I for different spatial lags k, the spatial correlogram, is useful for detecting scale variations in the spatial pattern (Upton and Fingleton).

Acknowledgement

None.

Conflicts of Interest

None.

References

1. Griffith DA (2003) Spatial Autocorrelation and Spatial Filtering: Gaining Through Theory and Scientific Visualization.
2. Tampa Hillsborough Homeless Initiative (2017) About Homelessness in Hillsborough County.
3. National Coalition for the Homeless (2007) Homeless Families with Children.
4. Crawford DM, Trotter EC, Sittner Hartshorn K, Whitbeck LB (2012) Pregnancy and Mental Health of Young Homeless Women.Am J Orthopsychiatry 81(2): 173-183.
5. Culhane D, Lee C, Wachter S (1996) Where the Homeless Come From: A Study of the Prior Address Distribution of Families Admitted to Public Shelters in New York City and Philadelphia. Fannie Mae Foundation 7(2): 327-364.
6. Burt M, Pearson C, Montgomery A (2007) Community-Wide Strategies for Preventing Homelessness: Recent Evidence. Journal of Primary Prevention (28): 213-228.
7. Baggett TP, O'Connell, Singer DE, Rigotti NA (2010) The Unmet Health Care Needs of Homeless Adults: A National Study. Am J Public Health 100(7): 1326-1333.
8. Hwang S W and Henderson J (2010) Health Care Utilization in Homeless People: Translating Research into Policy and Practice.
9. Moorkath F, Vranda M, Naveenkumar C (2018) Live without Roots: Institutionalized Homeless Women with Chronic Mental Illness. Indian J Psyho Med 40(5): 467-481.
10. National Alliance to End Homelessness (2020) The State of Homelessness in America.
11. National Coalition for the Homeless (2007) Domestic Violence and Homelessness.
12. National Conference of State Legislatures (2019) Youth Homelessness Overview.
13. Wen CK, Hudak LP, Hwang SW (2007) Homeless People's Perceptions of Welcome ness and Unwelcomeness in Healthcare Encounters. J Gen Intern Med 22(7): 1011-1017.