Spatial Statistics 45 (2021) 100523

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Addressing uncertainty in census estimates

Noah Lorincz-Comi [a,*], Jayakrishnan Ajayakumar [a], Jacqueline Curtis [a], Jing Zhang [a], Andrew Curtis [a], Rachel Lovell [b]

[a] *GIS Health & Hazards Lab, Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, United States of America*
[b] *Department of Criminology, Anthropology, and Sociology, Cleveland State University, Cleveland, OH, United States of America*

A R T I C L E I N F O

Article history:
Received 9 March 2021
Received in revised form 21 May 2021
Accepted 25 May 2021
Available online 31 May 2021

Keywords:
Census
Spatial
Statistics
Epidemiology
Social determinants
Measurement error

A B S T R A C T

Data from the American Community Survey (ACS) provides a wealth of information useful for learning more about social determinants of health and their spatial distributions within a defined region. Although available data includes quantified indicators of uncertainty in aggregated location-specific estimates for a range of variables, this uncertainty is often ignored, the consequences of which may include estimate bias and reduced statistical power. Fortunately, the measurement error literature provides a range of useful tools for handling such error. We propose and demonstrate a new application of existing, well-supported measurement error models to spatial regression models. We show that the existing solution of ignoring the measurement error inherent in these data precludes precise effect estimation and that straightforward modifications to traditional estimators can be made to correct for this error. We intend for this work to establish the basic principles of error correction in spatial data and a new method for applying corrected regression estimators to such data.

1. Introduction

The American Community Survey (ACS; US Census Bureau, 2020) is an abundant source of social, economic, and behavioral information for small- and large-scale geographical locations in the

* Corresponding author.
E-mail address: njl96@case.edu (N. Lorincz-Comi).

https://doi.org/10.1016/j.spasta.2021.100523

United States. The data produced by these surveys are publicly available and commonly used for better understanding the spatial distributions of sociodemographic characteristics (e.g., Pollard and Jacobsen, 2017) and their associations with a range of diseases and conditions (e.g., Buerhaus et al., 2020). These data are collected by randomly selecting households within defined sampling blocks for each county in the US. All individuals within each household may contribute data. Data collected from within households can then be aggregated to the census tract level by taking, for example, the mean of all household values with each tract. The census tract may then become the unit of observation in subsequent analyses. Tract-level mean estimates have associated standard errors representing sampling error (see US Census Bureau, 2020 for complete methodological information), and these standard errors are then used to evaluate uncertainty in the data.

This error is typically either ignored, acknowledged but uncorrected, reduced by additional spatial aggregation, or intended to be eliminated by removing observations for which the estimates are considerably uncertain as defined by National Research Council standards (Sun and Wong, 2010; Folch et al., 2016; Wong and Sun, 2013; Wei and Grubesic, 2017; NRC, 2007). None of these solutions are optimal as they may introduce bias, reduce the amount of available information, and/or reduce statistical power to detect existing effects. Further, more error has been introduced in recent years as the ACS target number of households sampled within each small-area location has been reduced from 1 in 6 to 1 in 40 (MacDonald, 2006). Another important aspect of this error is that it may cluster spatially in those areas typically associated with the highest risk of illness, such as low income, resource availability, or percentage minority ethnicity (Bazuin and Fraser, 2013; O'Hare, 2019).

We propose a method to correct for this ubiquitous error by referring to the well-supported principles established in the measurement error literature. Measurement error models are those that correct for introduced error, in the form of instrument or sampling error, to allow for unbiased estimates and inferences. We present a straightforward extension of these models to incorporate the use of spatially varying coefficients such as in geographically weighted regression (GWR). We demonstrate that uncontrolled error can produce bias and/or reduced power under several conditions commonly observed in spatial research but that corrected parameter estimators can address this nonignorable error.

2. Current problem and proposed solution

2.1. Current problem

For a univariable linear regression model, call X_i the true independent variable, x_i its realized value measured in sampling with random error u_i at a particular location in a spatial area of $i = 1, 2, \ldots, n$ locations. The basic additive measurement error model is $x_i = X_i + u_i$, where all variables are treated as random (see Fuller, 2009 for a similar model in which X_i and Y_i are treated as fixed). The corresponding model for the dependent variable is $y_i = Y_i + q_i$. We assume that $E(u_i|X_i) = E(q_i|Y_i) = 0$ and that these errors are normally distributed about X_i and Y_i, respectively (it may not be under all circumstances a requirement that the errors are distributed normally). As a result, $E(x_i) = E(X_i)$ and $E(y_i) = E(Y_i)$. For a multivariable analysis, the x_i and u_i values simply become matrices with corresponding elements. Measurement error can be introduced as instrument error and/or sampling error. An estimate of instrument error is absent in ACS census data, but an estimate of sampling error is present for each estimate for each location. Instrument error is introduced when the measurement tool (e.g., a survey) imperfectly measures the construct of interest. Sampling error is the error in estimates introduced by having access to only a subset of the full population within a location. Given the relatively simple nature of some of the collected variables, we may be willing to assume that instrument error of the ACS census survey is minimal for some variables. Where instrument error is present in addition to sampling error, the sampling error will only capture a proportion of the total measurement error. Although we do not know the actual u_i or q_i values, the sampling errors of x_i and y_i (the standard errors of the location-specific estimates, call σ_{x_i} and σ_{y_i}) can be used to approximate the u_i and q_i distributions within each location using the available margin of error (MoE) as $\hat{\sigma}_{ui} = MoE_{x_i}/1.645$, where $\hat{\sigma}_{ui}/\hat{x}_i$ (which is correspondingly the same for q_i and y_i) produces the popular coefficient of variation (CoV) value commonly used in spatial research for assessing location-specific estimate precision. *MoE* is available for all variables in all recent releases of ACS data.

The measurement error literature provides a comprehensive description of the issues inherent in ignored error and advantages of solutions intended to correct for such error (see Fuller, 2009; Buonaccorsi, 2010; Wiley and Wiley, 1970). We here aim to describe the bias introduced by ignoring known sampling error on either X_i or Y_i in a univariable geographically weighted regression (GWR) model (the multivariable extension is present in the Supplementary section S2). For easier demonstration and interpretation, we will initially present results for a hypothetical single location within a large-area geographical space, these results only changing between locations based on location-specific weighting schemes. If we make some reasonable assumptions about the distributions of (x_i, y_i) and (u_i, q_i), we can identify the exact bias introduced by ignoring sampling error for a model fit at a particular location using knowledge of the bivariate normal (X_i, Y_i) and (x_i, y_i) distributions. A naïve GWR estimator for a single location will estimate (β_0, β_1) from the realized (x_i, y_i) values where sampling error is unknown or ignored. The naïve estimators of β_1 and β_0 are thus $\hat{\gamma}_1 = \sigma_{xy}/\sigma_x^2$ and $\hat{\gamma}_0 = \bar{y} - \bar{x}\hat{\gamma}_1$, respectively, where σ_{xy} is the distance-weighted sample covariance between observed x_i and y_i measured with error, σ_x^2 is the corresponding distance-weighted variance of x_i, and $\bar{x}$ and $\bar{y}$ are distance-weighted sample means. We thus call $\left(\hat{\gamma}_0, \hat{\gamma}_1\right)$ the naïve, uncorrected geographically weighted regression (GWR) model-estimated effects for a particular location.

2.2. Proposed solution

Given the above measurement error model, we want to describe the bivariate distribution of (x_i, y_i) using knowledge of the corresponding distributions of the true population values and their sampling errors for a GWR model fit at a single location i. The bias introduced at a single location approximates the bias introduced at every other location since different locations receive different weightings for each n location-specific $\left(\sigma_x, \sigma_y, \sigma_u, \sigma_q\right)$ values. We assume that (X_i, Y_i) are normally distributed about (μ_X, μ_Y) with variances $\left(\sigma_X^2, \sigma_Y^2\right)$, (x_i, y_i) are distributed about (X_i, Y_i) with variances $\left(\sigma_u^2, \sigma_q^2\right)$, and by extension (x_i, y_i) are distributed about (μ_X, μ_Y) with variances $\left(\sigma_X^2 + \sigma_u^2 + 2\sigma_{Xu}, \sigma_Y^2 + \sigma_q^2 + 2\sigma_{Yq}\right)$ and covariances $\left(\sigma_{XY} + \sigma_{uq}, \sigma_{XY} + \sigma_{uq}\right)$. We call the variances of the sampling errors u_i and q_i the distance-weighted mean sampling error variances (i.e., the mean of the squared margins of error divided by 1.645^2) that are supplied in the original ACS data that for all locations (as in Buonaccorsi, 2010). Even with only the variances of the sampling errors supplied in ACS data releases, one may choose to use the sampling errors themselves (or a transformed version) to produce an estimate of the sampling error covariance matrix. Not all correction methods available in the literature allow nonzero σ_{Xu} and σ_{Yq} values (termed 'nonclassical measurement error'). As at least $\sigma_{Xu} \neq 0$ may be commonly observed in practice, we present a model allowing and correcting for this.

We may say that for a particular location $E\left(Y|X\right) = \beta_0 + \beta_1 X$ and $\mu_Y = \beta_0 + \beta_1\mu_X$, $\sigma_Y^2 = \beta_1^2\sigma_X^2 + \sigma_\epsilon^2$, $\sigma_{xy} = Cov\left(X + u, Y + q\right)$, and $\sigma_{XY} = \beta_1\sigma_X^2$. We can thus alternatively express the distribution of (x_i, y_i) as:

$$(x_i, y_i) \sim N\left(\begin{bmatrix} \mu_X \\ \beta_0 + \beta_1\mu_X \end{bmatrix}, \begin{bmatrix} \sigma_X^2 + \sigma_u^2 + 2\sigma_{Xu} & \beta_1\sigma_X^2 + \beta_1\sigma_{Xu} + \sigma_{uq} \\ \beta_1\sigma_X^2 + \beta_1\sigma_{Xu} + \sigma_{uq} & \beta_1^2\sigma_X^2 + \sigma_\epsilon^2 + \sigma_q^2 + 2\sigma_{Yq} \end{bmatrix}\right). \quad (1)$$

The bias of $\hat{\gamma}_1$ in estimating β_1 for a given single location can be identified from the bivariate (x_i, y_i) distribution using:

$$E(\hat{\gamma}_1) = \sigma_{xy}/\sigma_x^2 = \beta_1(\sigma_X^2 + \sigma_{Xu})\left(\sigma_X^2 + \sigma_u^2 + 2\sigma_{Xu}\right)^{-1} + \sigma_{uq}\left(\sigma_X^2 + \sigma_u^2 + 2\sigma_{Xu}\right)^{-1}. \quad (2)$$

where $\sigma_{Xu} = \sigma_{uq} = 0$, $\hat{\gamma}_1$ is biased downward by a factor of $\sigma_X^2\left(\sigma_X^2 + \sigma_u^2\right)^{-1} = \kappa$, or by $100\left(1 - \kappa\right)\%$. The corrected $\hat{\beta}_1$ estimate is produced by solving the $\hat{\gamma}_1$ equation for β_1 and using appropriate maximum likelihood moment estimators that, in GWR, are weighted by location. Variances for these

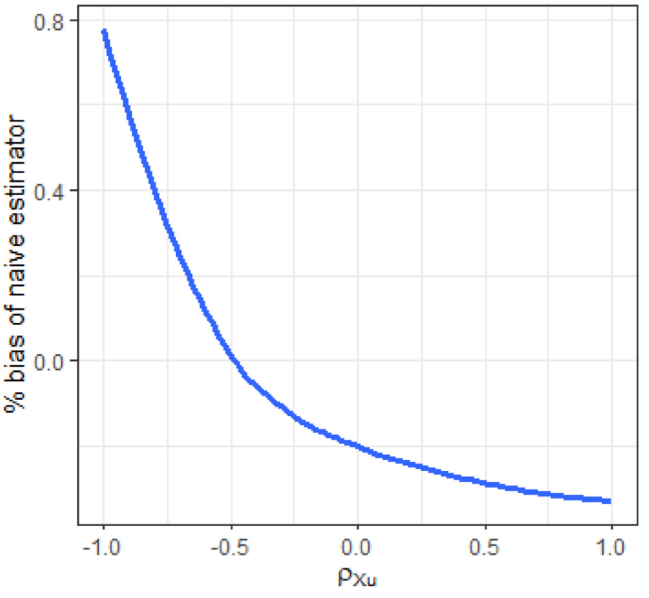

Fig. 1. Bias in naïve estimator at different error correlations for single location. This plot displays the % bias in the naïve estimator compared to the corrected estimator under different correlations between true X *values and their corresponding sampling errors.*

estimates, and those in the multivariable extension, are provided in the Supplementary sections S1 and S2, respectively. It can also be seen that the bias in $\hat{\gamma}_1$ can go in either direction depending on the value of σ_{Xu}. Fig. 1 displays the percentage bias in $\hat{\gamma}_1$ at ρ_{Xu} values ranging from $\{-1, 1\}$ when $\sigma_X^2 = \sigma_Y^2 = 1$, $\rho_{XY} = 0.75$, $\sigma_u^2 = 0.25$, and $y = Y$. These results demonstrate that the bias can be increased or decreased by the correlation of X_i with its sampling error but can also be eliminated under certain conditions.

Hypothesis testing of the naïve estimator $\hat{\gamma}_1$ (H_0: $\hat{\gamma}_1 = 0$) is still valid if $q_i = 0$ for $i = 1, 2, \ldots, n$, or if any $q_i \neq 0$ but $\sigma_{uq} = 0$. There is an inverse relationship between σ_{uq} and power of the corrected estimator. Fig. 2 displays how σ_{uq} affects Type I and II error rates of the naïve estimator at $\rho_{XY} = 0.35$, $n \in \{50, 100, 300\}$, $\alpha = 0.05$, and fixed $\sigma_q = \sigma_u = 0.25$ using simulation. These results indicate that the Type I error rate increases and the Type II error rate decreases as σ_{uq} becomes larger under the naïve estimator $\hat{\gamma}_1$. The Type I and Type II error rates are generally 0 for $\sigma_{uq} < 0$ and $\sigma_{uq} > 0$, respectively.

The above measurement error model and discussion can also be used to produce a corrected estimate of the correlation between X and Y (and thus $\hat{R}^2_{XY}$ in the univariable case):

$$\hat{\rho}_{XY} = \hat{\sigma}_{XY}\left(\hat{\sigma}_X\hat{\sigma}_Y\right)^{-1} = \left(\sigma_{xy} - \sigma_{uq}\right)\left[\left(\sigma_x^2 - \sigma_u^2 - 2\sigma_{xu}\right)\left(\sigma_y^2 - \sigma_q^2 - 2\sigma_{yq}\right)\right]^{-1/2} \tag{3}$$

where each variance/covariance is weighted according to a particular location. We may also extend this to produce a corrected Moran's I estimator of spatial autocorrelation for a single variable (Moran, 1950):

$$\hat{\rho}_I = n\hat{\sigma}_{XY}\left(\hat{\sigma}_x^2\sum_{i=1}^{n} w_i\right)^{-1} = n\left(\sigma_{xy} - \sigma_{uq}\right)\left[\left(\sigma_x^2 - \sigma_u^2 - 2\sigma_{xu}\right)\sum_{i=1}^{n} w_i\right]^{-1} \tag{4}$$

where the covariances for each variable are calculating by summing the weighted squared deviations from the respective weighted means over $i, j = 1, 2, \ldots, n$ for $i \neq j$. That is, the products of squared mean differences are calculated between locations.

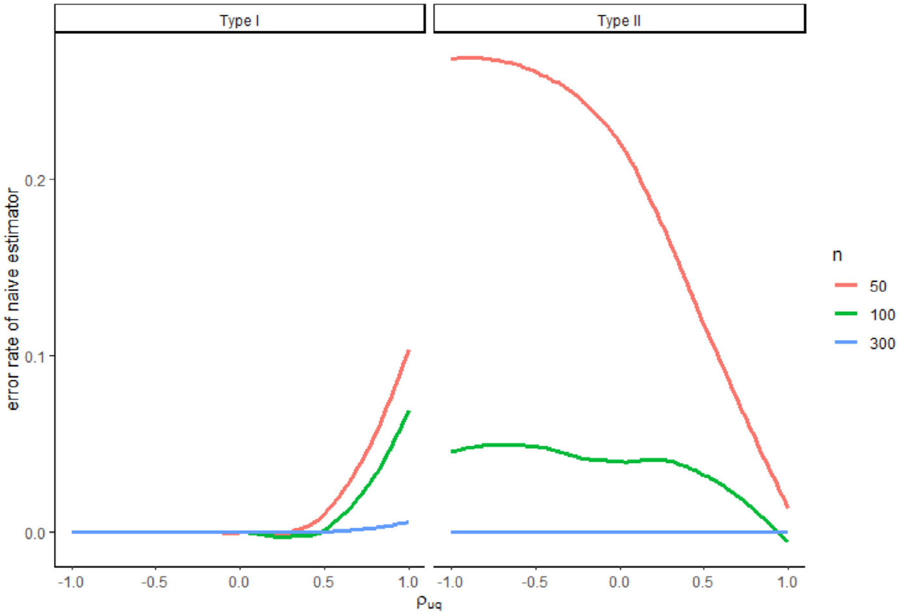

Fig. 2. Type I, II error rates at different (u, q) correlations. u and q are the measurement errors on true X *and* Y values, respectively. Type I and II error rates are determined using simulation at $n \in \{50, 100, 300\}$, $\sigma_u = \sigma_q = 0.25$, and $\rho_{XY} = 0.35$ and comparing the naïve estimator to the corrected estimator.

3. Bias analysis when sampling error clusters spatially

We next demonstrate the behavior of the naïve univariable estimator under different σ_{Xu} conditions where $\sigma_q^2 = 0$ for multiple locations to better understand how different spatial correlation structures of X_i with its sampling error produce different spatial distributions of bias. It is often the case in practice that ACS census data is used as the source for covariate but not outcome variables and thus, we may often be forced in practice to assume $y_i = Y_i$. Given a spatial area of 100 equally spaced locations separated from neighboring locations by a distance value of 1, we aim to demonstrate bias in the naïve estimator and the spatial distributions of bias under different ρ_{Xu} values using simulations. For these simulations, we define a population multivariate truncated normal distribution, draw $B = 1500$ samples from it and estimate $\left(\hat{\gamma}_{0i}, \hat{\gamma}_{1i}\right)$ and $\left(\hat{\beta}_{0i}, \hat{\beta}_{1i}\right)$ each time from the sample values for locations $i = 1, 2, \ldots, 100$. When fitting the naïve and corrected models, the moments of the sampling distributions are weighted using location-specific weights. Additionally, X_i and Y_i were made to be correlated with the weights vector at a particular location to create spatial autocorrelation of intensity $\tilde{\rho}_X = \tilde{\rho}_Y = 0.8$. The parameterization of the distributions from which the samples were drawn is described fully in Supplementary section S3. In brief, all parameters are fixed across the simulations but ρ_{Xu} takes on values of 0, −0.5, or 0.5. Under the latter conditions, since true X values are spatially autocorrelated, error u values also were. The variances of u and q are set to 0.25, one-fourth of the variances of X and Y.

The results of these simulations (as displayed in Fig. 3) demonstrate again that moderate positive correlation between true X values with their sampling errors can introduce additional bias beyond that which is already included by the sampling error alone. Although, under these simulation conditions, this additional bias is small. The largest bias values spatially clustered in the locations in which the true correlation between X and Y is strongest and low bias values where this correlation is weakest. Conversely, when the correlation between the true value and its sampling error is moderately strong and negative, a large proportion of the bias can be eliminated. The spatial

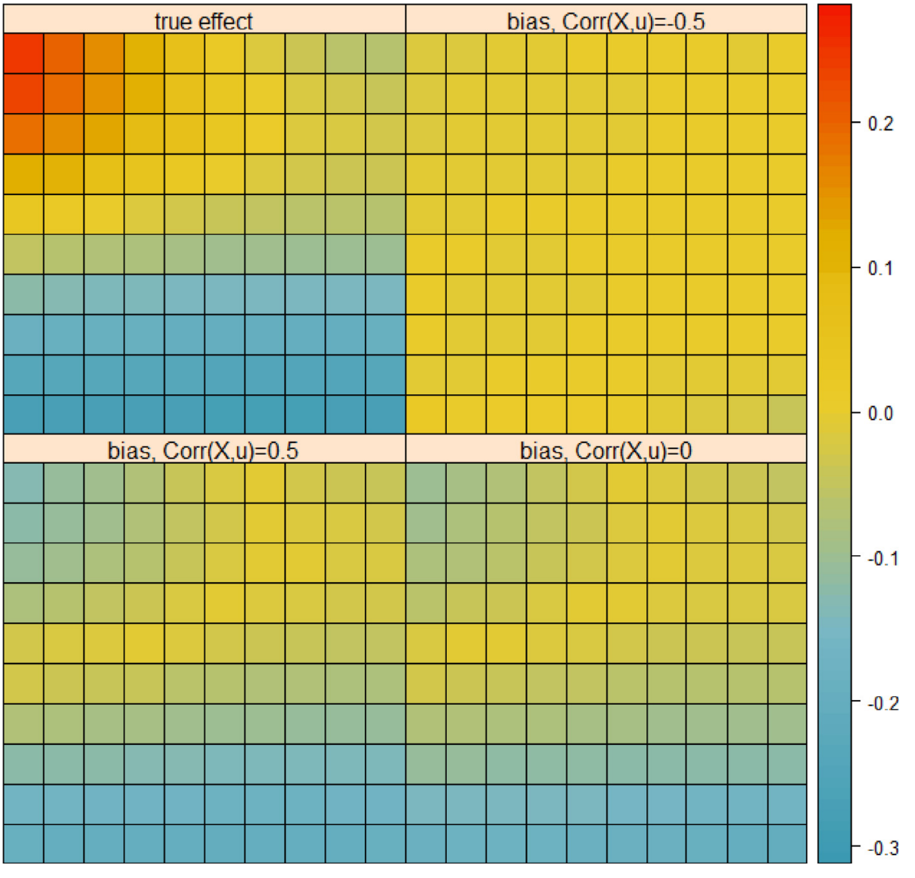

Fig. 3. Spatial distributions of bias under different (X, u) correlations. Plotted in the upper left panel is the true estimate of β_1 for each location when $x_i = X_i, y_i = Y_i$. Plotted in the remaining three panels are bias values (equal to the difference $\left|\hat{\gamma}_{1i}\right| - \left|\hat{\beta}_{1i}\right|$) for each location when $\rho_{Xu} = 0$, $\rho_{Xu} = 0.5$, or $\rho_{Xu} = -0.5$. Warmer colors (yellows, reds) indicate larger values; cooler colors (blues) indicate smaller values . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

distribution of bias is similar between moderate positive and zero correlations between X and its sampling error. The stationary distributions of bias also differ between these conditions. Compared to the $\rho_{Xu} = 0$ condition, under $\rho_{Xu} = 0.5$ there is similar variability (SD: 0.06 vs. 0.06, respectively) and a similar range (0.21 vs. 0.20) of bias, but under $\rho_{Xu} = -0.5$ there is less variability (0.06 vs. 0.01) and a smaller range (0.21, 0.06). Compared to $\rho_{Xu} = 0.5$, the $\rho_{Xu} = -0.5$ condition produced bias of less variation and smaller range. Table 1 contains these results.

Lastly, for a fixed WLS $\hat{R}^2_{XY}$ under $\sigma_{Xu} = \sigma_{uq} = 0$ (i.e., the bias in $\hat{\gamma}_1$ is determined by κ), we can determine at what sample size the weighted mean squared error of the corrected estimator at a single location is less than the corresponding error of the naïve estimator using the following formula:

$$n_{min} = 3 - \frac{\left(-R^2_{XY} + 1\right)(\kappa + 1)}{R^2_{XY}\,(\kappa - 1)} \tag{5}$$

where κ is defined above. This equation can be applied to all locations in a defined spatial area. The derivation of this equation is presented in the Supplementary section S4. Fig. 4 displays these

Table 1
Distributions of bias under different sampling error correlation structures.

	Mean	SD	Min	Max
Corr(X, u) = 0.5	−0.09	0.06	−0.20	−1.1e−3
Corr(X, u) = 0	−0.08	0.06	−0.21	−1.7e−3
Corr(X, u) = −0.5	−7.3e−4	0.01	−0.04	0.02

These values are produced under a geographically weighted regression (GWR) model when both true X and Y are positively spatially correlated with spatial autocorrelation values $\tilde{\rho}_X = \tilde{\rho}_Y = 0.8$ across a spatial grid of 100 equally spaced locations. Weights were assigned to each location using a Gaussian kernel: $w_i = \exp\left[-0.5\left(d_{ij}/2.5\right)^2\right]$ for $i, j = 1, 2, \ldots, n,\ i \neq j$. Sample $(x_i, y_i, u_i, q_i, w_i)$ values were drawn 1000 times from a truncated multivariate normal distribution with all parameters of 0 value except $\mu_X = \mu_Y = \sigma_X^2 = \sigma_Y^2 = 1$, $\sigma_u^2 = \sigma_q^2 = 0.25$, $\sigma_{XY} = 0.5$, $\sigma_w^2 = 0.06$, $\mu_w = 0.13$, and $\rho_{Xu} = \{-0.5, 0, 0.5\}$ and fit to the naïve and corrected models to calculate bias values. Weights were included in this multivariate distribution as a random variable to set the spatial autocorrelation. The realization of weight values w_i from the multivariate distribution was not used in fitting the naïve and corrected models. Instead, the fixed Gaussian kernel distance-based weights were used. The distribution is only truncated to ensure all weights were greater than 0 and less than or equal to 1.

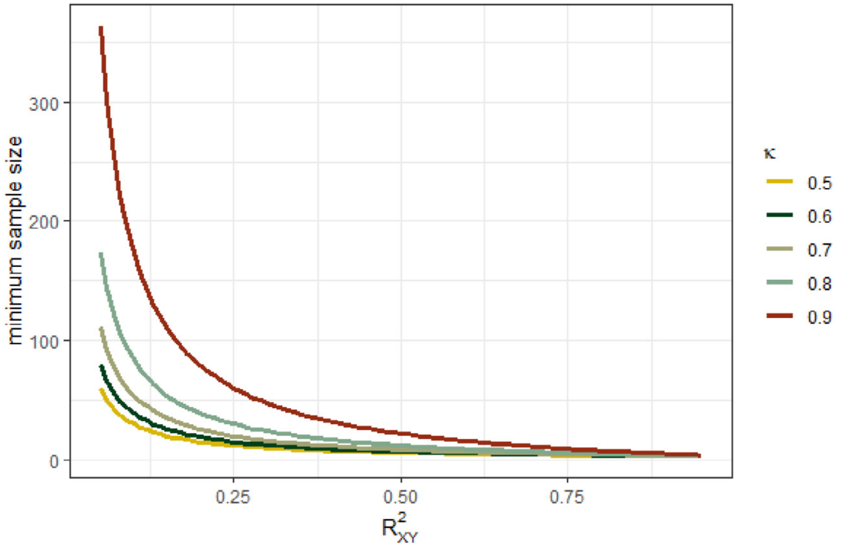

Fig. 4. Minimum sample sizes at which naïve model error exceeds corrected model error. These values are generated by solving an equation. This equation sets the mean squared error of the naïve estimator equal to the mean squared error of the corrected estimator. Minimum sample sizes are produced by solving this equation for n (see the text for solved version and Supplementary section S4 for solving of extended version)

minimum sample sizes for WLS R^2_{XY} values of 0 to 1 and κ values of $0.5, 0.6, \ldots, 0.9$. These results indicate that it is advantageous to use the corrected estimator even at relatively small sample sizes (<100) when the correlation between the independent and dependent variables at least mild (>0.2) and κ is <0.9.

4. Discussion

ACS census data is a valuable tool for learning about social determinants of health and their spatial distributions. However, this data contains error that is routinely ignored in practice. We have presented the problems associated with ignoring this error and proposed a general correction that may make estimated associations observed in spatial data more precise and inferences more accurate. Notably, if the true covariate values are negatively correlated with their sampling errors, the bias introduced by a naïve estimator may be minimal (although not uniformly spatially distributed). However, if this correlation is positive, the bias can be increased further, especially in those locations where the association between the independent and dependent variable is strongest.

4.1. Limitations

The added complexity of the univariable, and especially the multivariable, corrected models may serve as a deterrent from application in practice. Additionally, we have omitted from this text mention of the variances of the corrected estimators. The parameter variance estimators in the measurement error and GWR literature cannot be applied to a model that incorporates both aspects of sampling error and spatially varying coefficients. We thus propose a jackknife approach in the Supplementary section A2 that may be used in either the univariable or multivariable cases.

4.2. Strengths

Our proposed correction is strengthened by the comprehensive support of the measurement error literature (e.g., Fuller, 2009; Buonaccorsi, 2010; Wiley and Wiley, 1970). The presented corrective methods are relatively easy to implement and congruent to standard models in their interpretation. In addition to the presented model, a short list of alternative models is also available (e.g., instrumental variable estimation, replication data, or SIMEX such as in Alexeeff et al., 2016) that may be more appropriate for a particular research question. Lastly, we have begun developing software (found at github.com/noahlorinczcomi/spatialme) to fit these models and their multivariable extensions to spatial data using R.

5. Conclusion

Treating ACS data as measured without error can produce biased analyses and subsequently incorrect epidemiological inferences, especially in the locations in which disease risk is potentially the greatest. The adjustment method we propose here, a well-established statistical tool previously popular in non-spatial settings, is straightforwardly applied to spatial data. The application of this proposed method to spatial settings can expand the spatial epidemiologist's methodological toolkit and may allow for more meaningful insights into spatial distributions of disease risk. This becomes most important for the locations in which disease risk and burden are the greatest. It is routinely in these locations where ACS sampling error has the greatest negative impact on statistical analyses. It is our hope that more research in this area of error correction is conducted and the optimal solutions for a range of unique conditions identified. After such work, we may begin to more fully understand how spatial distributions of social/economic factors are related to health and as such accumulate the information necessary to intervene effectively in places where it is most urgently required.

CRediT authorship contribution statement

Noah Lorincz-Comi: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - original draft. **Jayakrishnan Ajayakumar:** Conceptualization, Methodology, Software, Writing - original draft, Supervision. **Jacqueline Curtis:** Conceptualization, Writing - original draft, Supervision. **Jing Zhang:** Conceptualization, Methodology, Visualization. **Andrew Curtis:** Conceptualization, Writing - original draft, Supervision. **Rachel Lovell:** Writing - original draft, Supervision.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Additional derivations, such as estimator variances, multivariable extensions, and simulation data generators can be found online at https://doi.org/10.1016/j.spasta.2021.100523.

References

Alexeeff, S.E., Carroll, R.J., Coull, B., 2016. Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. Biostatistics 17 (2), 377–389.

Bazuin, J.T., Fraser, J.C., 2013. How the ACS gets it wrong: The story of the American Community Survey and a small, inner city neighborhood. Appl. Geogr. 45, 292–302.

Buerhaus, P.I., Auerbach, D.I., Staiger, D.O., 2020. Older clinicians and the surge in novel coronavirus disease 2019 (COVID-19). JAMA 323 (18), 1777–1778.

Buonaccorsi, J.P., 2010. Measurement Error: Models, Methods, and Applications. CRC press.

Folch, D.C., Arribas-Bel, D., Koschinsky, J., Spielman, S.E., 2016. Spatial variation in the quality of American Community Survey estimates. Demography 53 (5), 1535–1554.

Fuller, W.A., 2009. Measurement Error Models, Vol. 305. John Wiley & Sons.

MacDonald, H., 2006. The American Community Survey: Warmer (more current), but fuzzier (less precise) than the decennial census. J. Am. Plan. Assoc. 72 (4), 491–503.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1), 17–23. http://dx.doi.org/10.2307/2332142, JSTOR 2332142.

National Research Council, 2007. In: Citro, Constance F., Kalton, Graham (Eds.), Using the American Community Survey: Benefits and Challenges. National Academies Press.

O'Hare, W.P., 2019. Differential Undercounts in the US Census: Who is Missed? Springer Nature.

Pollard, K., Jacobsen, L.A., 2017. The Appalachian Region: A Data Overview from the 2011-2015 American Community Survey. Chartbook. Appalachian Regional Commission.

Sun, M., Wong, D.W., 2010. Incorporating data quality information in mapping American Community Survey data. Cartogr. Geogr. Inf. Sci. 37 (4), 285–299.

US Census Bureau, 2020. American Community Survey: Design and Methodology Report. URL: https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html.

Wei, R., Grubesic, T.H., 2017. An alternative classification scheme for uncertain attribute mapping. Prof. Geogr. 69 (4), 604–615.

Wiley, D.E., Wiley, J.A., 1970. The estimation of measurement error in panel data. Am. Sociol. Rev. 35 (1), 112–117.

Wong, D.W., Sun, M., 2013. Handling data quality information of survey data in GIS: A case of using the American Community Survey data. Spat. Demogr. 1 (1), 3–16.