Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 74(3) 2020 195–200

Prediction of Finite Population Total for Geo-referenced Data

Samir Barman, Pradip Basak and Hukum Chandra

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 14 January 2020; Revised 28 July 2020; Accepted 04 October 2020

SUMMARY

In many surveys (for example, agriculture, forestry, environmental and ecological surveys), data are spatially correlated and independence assumption is questionable. As a result, the existing estimators for population total (or mean) based on standard survey estimation method can be biased and less efficient. Use of spatial information in sample surveys is expected to provide a better estimation of population parameters. This paper develops the estimators for finite population total incorporating spatial information. The proposed estimators are evaluated through simulation studies. The empirical results show that the developed estimators have smaller bias and better efficiency as compared to the existing estimators.

Keywords: Geo-referenced data, Population total, Spatial information.

1. INTRODUCTION

Now a days, various modern technologies such as Computer Assisted Personal Interviewing (CAPI), Global Positioning System (GPS) instrument etc. are widely used in different surveys for data collection. The technologies make it easy to collect geo-referenced information. Hence, geo-referenced data is now readily available. This type of data describes the objects in terms of their position with a known coordinate system, their attributes and spatial interactions with each other. It is important to note that existing survey estimation methods rarely make use of such enriched geographical information. For example, under classical linear model based estimation of finite population total, population units are assumed to be independent. In many surveys (e.g., agricultural and environmental surveys), data are often spatially correlated. Therefore, assumption of independence is often violated and it also leads to biased and less efficient estimate of the parameters. However, to the best of our knowledge, very limited work has been undertaken to incorporate the geo-referenced information for the estimation of finite population total or mean.

In literature, several approaches are available to model the geo-referenced data. In last few decades, Geographically Weighted Regression (GWR) method has emerged as one of the popular approach to model geo-referenced data. In GWR, model parameters are estimated location-wise over the study space (Brunsdon *et al.*, 1996 and 1998). Fotheringham *et al.* (1998) used two statistical techniques that are GWR and expansion method to examine the spatial variability of regression results across a region and so inform on the presence of spatial non-stationarity. Leung *et al.* (2000) studied a prediction problem for the analysis of spatial non-stationarity under GWR approach and also developed a statistical method for testing the goodness of fit of the GWR model that made it possible to test spatial non-stationarity in a conventional statistical manner. GWR model is mainly used for prediction of spatial characteristics rather than inference (Montanari *et al.*, 2010). Chandra *et al.* (2012) developed linear mixed model version of GWR and applied this model in small area estimation of means. Chandra *et al.* (2017) considered the problem of fixed effect parameter estimation of generalized linear mixed model (GLMM)

Corresponding author: Pradip Basak
E-mail address: pradipbasak.99@gmail.com

at the presence of spatial non-stationarity in the population under small area estimation.

GWR leads to a location specific model. In model based survey estimation, GWR approach can be used to model the geo-referenced data. The use of spatial information via GWR model is expected to provide a better estimation of population parameters. So there is a growing interest in development of methods that use geo-reference data for predicting the finite population total. It is therefore timely and important to explore the use of geo-referenced data for the estimation of finite population total. In section 2, the traditional estimators for prediction of finite population total are described. Section 3 presents the proposed estimators of finite population total based on GWR approach. In section 4, results of the empirical evaluations are described. Finally section 5 gives the concluding remarks.

2. PREDICTORS FOR FINITE POPULATION TOTAL

To start, let us consider a finite population $U=\{1,2,...,N\}$ of N units such that each unit of population are indexed by 'i'. Let $\mathbf{y}=(y_1,y_2,...,y_N)$ denotes the study variable, and $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,...\mathbf{x}_N)^T$ a set of $N\times p$ of auxiliary variables where each $\mathbf{x}_i=(x_{i1},...,x_{ip})^T$, $\forall i\in U$, is a p-auxiliary variable associated with each study variable. Also let $L=(l_1,l_2,...,l_N)$ be the vector of location of population units where $l_i=(\text{lon},\text{lat})$ denotes the geographic location of i^{th} unit. It is assumed that the population totals of auxiliary variables are known and there exist a linear relationship between study variable and auxiliary variables. Let s be a sample of size n drawn from this population. Without losing any information, the population can be grouped into sample and non-sample part, of size n and $N-n$ respectively. The non-sample part is denoted by r. Under the model based approach (Valliant *et al.* 2000), let us consider a linear model as

$$y_i=\mathbf{x}_i^T\boldsymbol{\beta}+e_i \tag{1}$$

where $\boldsymbol{\beta}$ is the unknown regression coefficient and e_i is the random error component identically and independently distributed as normal with mean zero and variance $\sigma^2\forall i=1,...,N$. The above model at the population level can be expressed in matrix form as

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{e} \tag{2}$$

where $\mathbf{y}$ and $\mathbf{e}$ are of order $N\times 1$, $\boldsymbol{\beta}$ is of order $p\times 1$ and $\mathbf{X}$ is of order $N\times p$. Also $E(\mathbf{e})=\mathbf{0}$ and $V(\mathbf{e})=\mathbf{V}=\sigma^2\mathbf{I}_N$. Based on this, our interest is to find the population total. The sample and non-sample partition of **y, X** and **V** is given by

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_s\\ \mathbf{y}_r\end{bmatrix},\ \mathbf{X}=\begin{bmatrix}\mathbf{X}_s\\ \mathbf{X}_r\end{bmatrix}\text{ and }\mathbf{V}=\begin{bmatrix}\mathbf{V}_{ss} & \mathbf{V}_{sr}\\ \mathbf{V}_{rs} & \mathbf{V}_{rr}\end{bmatrix}.$$

The best linear unbiased estimator of regression parameter $\boldsymbol{\beta}$ under model (2) is $\hat{\boldsymbol{\beta}}=\left(\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}\mathbf{y}_s=\mathbf{H}_s\mathbf{y}_s$, where $\mathbf{H}_s=\mathbf{A}^{-1}\mathbf{X}_s^T\mathbf{V}_{ss}$ and $\mathbf{A}=\mathbf{X}_s^T\mathbf{V}_{ss}^{-1}\mathbf{X}_s$. The empirical best linear unbiased predictor (EBLUP) weight of population total is given as $\mathbf{g}=\mathbf{1}_n+\mathbf{H}_s^T\left(\mathbf{X}_N^T\mathbf{1}_N-\mathbf{X}_s^T\mathbf{1}_n\right)$.

The standard survey weighted estimator of finite population total is defined as

$$\hat{T}^{(1)}\quad\sum w_iy_i \tag{3}$$

where w_i is the survey weight.

Under the linear prediction model (2), an estimator of population total based on the predicted values of y can be obtained as

$$\hat{T}^{(2)}=\sum_{i=1}^{n}w_i\hat{y}_i=\sum_{i=1}^{n}w_i\mathbf{x}_i^T\hat{\boldsymbol{\beta}} \tag{4}$$

where $\hat{y}_i=\mathbf{x}_i^T\hat{\boldsymbol{\beta}}$ is the predicted value of y for i^{th} unit. However, an important drawback of this estimator is that it is based on the predicted value of y. Under SRSWOR, when there is no auxiliary information, $\hat{y}_i=\hat{\beta}_0=\bar{y},\forall i\in s$ and the corresponding estimator is simplified as $\hat{T}^{(2)}=\sum_{i=1}^{n}w_i\hat{y}_i=N\bar{y}$.

Under model based approach, an EBLUP of the finite population total is defined as

$$\hat{T}^{(3)}=\mathbf{g}^T\mathbf{y}_s \tag{5}$$

where $\mathbf{g}=(g_i;i=1,...,n)$ and g_i is the EBLUP weight of the i^{th} unit. Under SRSWOR, when there is no auxiliary information, $\mathbf{g}=1_n+\frac{N-n}{n}1_n$ and the corresponding estimator is simplified as $\hat{T}^{(3)}=\mathbf{g}^T\mathbf{y}_s=N\bar{y}$. It is clear that under SRSWOR when no auxiliary information is available all the existing estimators (*i.e.* $\hat{T}^{(1)}=\hat{T}^{(2)}=\hat{T}^{(3)}$) are identical.

3. PROPOSED ESTIMATORS FOR PREDICTION OF FINITE POPULATION TOTAL

Let us assume a location specific (GWR) model as

$$y_i(l_i) = \mathbf{x}_i^T \boldsymbol{\beta}_i + \varepsilon_i \; ; \; i = 1, 2, ..., N \tag{6}$$

where $\boldsymbol{\beta}_i = \boldsymbol{\beta}(l_i)$ is a vector of regression coefficient of order $p \times 1$ associated with location l_i, $\varepsilon_i \sim N(0, \sigma^2)$ and $\mathbf{x}_i$ is a vector of order of $p \times 1$ auxiliary information associated with y_i. At population level, model (6) can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}(L) + \mathbf{e} \tag{7}$$

Under GWR, the data are assumed to follow location specific regression function with geographical weight being used for estimation of parameters of this location specific or local regression function. The parameters are estimated using weighted least squares (WLS) with the weight varies location to location. Then, based on the observed sample, GWR method is used by assigning the weight to obtain the geographically weighted best linear unbiased estimator (GWBLUE) of $(\;)$ at location l_i as in the form

$$\hat{\boldsymbol{\beta}}(l_i) = \left(\mathbf{X}_s^T \boldsymbol{\Omega}_{ss}^{-1}(l_i)\mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \boldsymbol{\Omega}_{ss}^{-1}(l_i)\mathbf{y}_s = \mathbf{H}_s^{GWR}(l_i)\mathbf{y}_s \tag{8}$$

where $\boldsymbol{\Omega}_{ss}(l_i) = \mathbf{U}_s^{-1}(l_i)\mathbf{V}_{ss} = \sigma^2 \mathbf{U}_s^{-1}(l_i)$ and $\mathbf{H}_s^{GWR}(l_i) = \left(\mathbf{X}_s^T \mathbf{U}_s(l_i)\mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{U}_s(l_i)$.

Here $\mathbf{U}_s(l_i)$ is the matrix of weights called spatial weight matrix that is specified to the location l_i such that observations nearest to location l_i are given greater weight than the observations far away. This matrix is defined as

$$\mathbf{U}_s(l_i) = \begin{bmatrix} u_{i1} & 0 & \cdots & 0 \\ 0 & u_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{in} \end{bmatrix} \tag{9}$$

where u_{ij} is the spatial weight function based on Euclidian distance, gives weight to j^{th} unit with respect to i^{th} unit, is defined as

$$u_{ij} = \frac{1}{1 + d_{ij}} \tag{10}$$

where d_{ij} is the Euclidian distance between i^{th} and j^{th} unit. This function gives greater weight to the neatest observations than the observations that are far away. On substituting the value of u_{ij} obtained in (10) in the weight matrix (9), the weight matrix can be written as $\mathbf{U}_s(l_i)$. The corresponding WLS estimate of the model parameters is given by

$$\hat{\boldsymbol{\beta}}(l_i) = \left(\mathbf{X}_s^T \mathbf{U}_s(l_i)\mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \mathbf{U}_s(l_i)\mathbf{y}_s \tag{11}$$

An estimator of population total based on GWR approach is defined as

$$\hat{T}^{(4)} = \sum_{i=1}^{n} w_i \hat{y}_i(l_i) = \sum_{i=1}^{n} w_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(l_i) \tag{12}$$

where $\hat{y}_i(l_i) = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(l_i)$ is the predicted value of the study variable $y_i(l_i)$. This estimator uses the spatial information. Under SRSWOR, when there is no auxiliary information, the corresponding parameter $\boldsymbol{\beta}$ at location l_i become $\hat{\boldsymbol{\beta}}(l_i) = \hat{\beta}_0(l_i) = \frac{1}{D_i}\sum_{j=1}^{n} y_j u_{ij}$, where $D_i = \sum_{j=1}^{n} u_{ij}$ and the predicted value of $y_i(l_i)$ is given by $\hat{y}_i(l_i) = \hat{\beta}_0(l_i) = \frac{1}{D_i}\sum_{j=1}^{n} y_j u_{ij}$. The corresponding estimator becomes $\hat{T}^{(4)} = \sum_{i=1}^{n} w_i \hat{y}_i(l_i) = \sum_{i=1}^{n} w_i \left(\frac{1}{D_i}\sum_{j=1}^{n} y_j u_{ij}\right)$.

Another survey estimation technique has been developed by incorporating the spatial relationship of the variables under study in the linear regression model. Let us consider a statistical linear model that able to take into the account spatial varying behaviour as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{13}$$

where $\mathbf{y}$ is the dependent data vector of order $N \times 1$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, $\mathbf{X}$ is a covariate matrix of order $N \times p$ which contain spatial information and $\mathbf{e}$ is $N \times 1$ model error follows $N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ such that $E(\mathbf{e}) = 0$ and $V(\mathbf{e}) = \mathbf{V} = \sigma^2 \mathbf{I}_N$. Based on the sample, model parameters are estimated. Here we can't directly apply the ordinary least square method as a technique of parameter estimation. In this case we have used WLS method to estimate the unknown parameters of the model. The weights of WLS method represent the information about the population spatial variations. The weight matrix is a square symmetric matrix where individual weights are generated based on some distance function, defined as

$$\mathbf{U}_s^S = \begin{bmatrix} u_{11} & u_{21} & \ldots & u_{n1} \\ u_{12} & u_{22} & \ldots & u_{n2} \\ \vdots & \vdots & \ldots & \vdots \\ u_{1i} & u_{2i} & .. & u_{ni} \\ \vdots & \vdots & \ldots & \vdots \\ u_{1n} & u_{2n} & \ldots & u_{nn} \end{bmatrix} \quad (14)$$

where u_{ij} is the spatial weight function given to j^{th} location (j=1,…,n) in the sample to i^{th} (i=1,…,n) point based on the Euclidian distance. The corresponding WLS estimate of the model parameter can be denoted by $\hat{\boldsymbol{\beta}}^S = \mathbf{H}^S \mathbf{y}_s$ where $\mathbf{H}^S = \left(\mathbf{X}_s^T \left(\boldsymbol{\Omega}_{ss}^S\right)^{-1} \mathbf{X}_s\right)^{-1} \mathbf{X}_s^T \left(\boldsymbol{\Omega}_{ss}^S\right)^{-1}$ and $\boldsymbol{\Omega}_{ss}^S = \left(\mathbf{U}_s^S\right)^{-1} \mathbf{V}_{ss} = \sigma^2 \left(\mathbf{U}_s^S\right)^{-1}$. Under model assumption, the empirical best linear unbiased predictor (EBLUP) type weight of population total is given as

$$\mathbf{g}_s^S = \mathbf{1}_n + \left(\mathbf{H}_s^S\right)^T \left(\mathbf{X}_N^T \mathbf{1}_N - \mathbf{X}_s^T \mathbf{1}_n\right). \quad (15)$$

We have proposed an estimator that carries the spatial information of the variable based on the EBLUP type weight as

$$\hat{T}^{(5)} = \left(\mathbf{g}_s^S\right)^T \mathbf{y}_s = \sum_{i=1}^{n} g_i^S y_i . \quad (16)$$

where g_i^S is the EBLUP type weight of i^{th} unit. This estimator carries the spatial information through EBLUP type weight. Under SRSWOR, when there is no auxiliary information, the EBLUP type weight of the population total is then $\mathbf{g}_s^S = \mathbf{1}_n + \frac{N-n}{D}\left(D_1, D_2, ..., D_n\right)^T$, where $D = \sum_{i=1}^{n} D_i = \sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}$ and the estimator (16) leads to $\hat{T}^{(5)} = \sum_{i=1}^{n} y_i + \frac{N-n}{D}\sum_{i=1}^{n} D_i y_i$.

4. EMPIRICAL EVALUATIONS

Empirical evaluations have been carried out to investigate the finite sample performance of the proposed estimators for prediction of finite population total using geo-referenced data. The performance of the proposed estimators defined by $\hat{T}^{(4)}$ and $\hat{T}^{(5)}$ are compared with various existing estimators defined by $\hat{T}^{(1)}$, $\hat{T}^{(2)}$ and $\hat{T}^{(3)}$. In the simulation study, spatial population data is generated under a spatial dependent linear model with simultaneous auto regressive (SAR) type error structure, *i.e.*, effects of neighboring units have a SAR type correlation. Here, population size is considered as N=8100 and assumed to be located on a $\sqrt{N} \times \sqrt{N}$ grid with intersections uniformly spread between -1 to 1 and with a distance $2/\left(\sqrt{N}-1\right)$ between any two neighboring intersections along both the horizontal and vertical axes. Population data is generated under (i) no auxiliary variable: the model, $y_i = 50 + \alpha_i; i = 1,....,N$ and (ii) single auxiliary variable: the model, $y_i = 50 + x_i + \alpha_i; i = 1,....,N$, where the vector $\boldsymbol{\alpha} = (\alpha_i,\ i = 1,2,...N)$ of random errors are generated via a random draw from $N\left(\mathbf{0}, \sigma^2[(\mathbf{I}_N - \rho\mathbf{U})(\mathbf{I}_N - \rho(\mathbf{U})^T)]^{-1}\right)$, with values of ρ and σ set to 0.80 and 20 respectively. The auxiliary variable x_i is generated by chi-square distribution with 20 degrees of freedom. From this population, different samples of size, n=200 and 300 are selected using SRSWOR and the population total is estimated using different estimators included in the simulation study. The simulation has been run R=1500 times. The estimators are evaluated based upon the criteria of percentage absolute relative bias (ARB) and percentage relative root mean squared error (RRMSE), defined by

$$ARB = \frac{1}{R}\sum_{r=1}^{R}\left|\frac{\hat{T}_r - T}{T}\right| \times 100$$

$$RRMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{T}_r - T}{T}\right)^2} \times 100 ,$$

where T is the actual value of the population total, $\hat{T}_r$ is the predicted value of the population total for the r^{th} simulation run and R is the total number of simulation run. The results of the simulation study are presented in the following tables.

Table 1 presents the values of percentage absolute relative biases and percentage relative root mean squared errors recorded by the different estimators with auxiliary variable investigated in our simulations. These results show that both relative biases and relative root mean squared errors decreases with sample size of all the estimators. The proposed estimator $\hat{T}^{(4)}$ has the minimum relative biases as compared to all other estimators considered in the simulation study. The relative root mean squared errors of both the proposed estimators $\hat{T}^{(4)}$ and $\hat{T}^{(5)}$ are smaller than the existing estimators $\hat{T}^{(1)}$, $\hat{T}^{(2)}$ and $\hat{T}^{(3)}$. Further, between two proposed estimators, the estimator $\hat{T}^{(4)}$

is more efficient than the estimator $\hat{T}^{(5)}$. Overall, the proposed estimators outperform the existing estimators for spatial population.

The values of percentage absolute relative biases and percentage relative root mean squared errors of the different estimators with no auxiliary variable investigated in the simulations are set out in Table 2. Two points standout from these results. First, all the existing estimators are identical. Second, the proposed estimators show superior performance as compared to the existing estimators. In case of no auxiliary variable, between two proposed estimators, the estimator $\hat{T}^{(5)}$ is marginally better than the $\hat{T}^{(4)}$.

It worth noting that the proposed estimator $\hat{T}^{(4)}$ and $\hat{T}^{(5)}$ mainly differ in terms of the spatial weight matrices. These spatial weight matrices are used in estimating the model parameters. The estimator $\hat{T}^{(4)}$ is based on GWR approach which uses location specific spatial weight matrix to obtain location specific estimates of model parameters. On the other hand, in the estimator $\hat{T}^{(5)}$ a spatial weight matrix is used to incorporate spatial correlation in estimating the global estimates of model parameters. The performance of two estimators $\hat{T}^{(4)}$ and $\hat{T}^{(5)}$ in the simulation studies are essentially as one would expect. In GWR approach the regression coefficients associated with the auxiliary variables are location specific, so the estimator $\hat{T}^{(4)}$ dominates the estimator $\hat{T}^{(5)}$ when auxiliary variable is used. In case no auxiliary variable is used, the estimator $\hat{T}^{(5)}$ which uses the global estimates of parameters and incorporate spatial correlation outperforms the estimator $\hat{T}^{(4)}$.

5. CONCLUSIONS

This paper describes estimation of finite population total for geo-reference data. In particular, two different estimators are proposed for finite population total suitable for spatial population. The first estimator is developed expanding the GWR approach and using location specific spatial weight matrix whereas the second the estimator is motivated using a global spatial weight matrix defined for sample data. The empirical results based on simulation studies indicate that the proposed estimators perform better than existing estimators both in the presence and absence of auxiliary information. Hence, the proposed estimators for finite population total can be a sound alternative to existing estimators for spatial population data. Further, in this paper we have explored one type of spatial weight function based on inverse of Euclidian distance, thus there is a scope for improvement by exploring some other form of spatial weight function.

ACKNOWLEDGMENT

The authors would like to acknowledge the valuable comments and suggestions provided by the anonymous referee.

Table 1. Percentage absolute relative bias (ARB, %) and percentage relative root mean square error (RRMSE, %) of different estimators with auxiliary variable

Sample Size	Performance Criteria	Existing			Proposed		% Gain	
		$\hat{T}^{(1)}$	$\hat{T}^{(2)}$	$\hat{T}^{(3)}$	$\hat{T}^{(4)}$	$\hat{T}^{(5)}$	$\hat{T}^{(4)}$	$\hat{T}^{(5)}$
200	ARB	18.53	18.53	18.57	18.16	18.55	2.03	0.12
	RRMSE	24.28	24.28	24.29	23.71	24.03	2.21	0.23
300	AB	16.02	16.02	16.06	15.61	15.84	2.54	1.14
	RRMSE	20.49	20.49	20.54	19.94	20.37	2.69	0.85

Table 2. Percentage absolute relative bias (ARB, %) and percentage relative root mean square error (RRMSE, %) of different estimators with no auxiliary variable

Sample Size	Performance Criteria	Existing			Proposed		% Gain	
		$\hat{T}^{(1)}$	$\hat{T}^{(2)}$	$\hat{T}^{(3)}$	$\hat{T}^{(4)}$	$\hat{T}^{(5)}$	$\hat{T}^{(4)}$	$\hat{T}^{(5)}$
200	ARB	26.00	26.00	26.00	25.59	25.57	1.58	1.63
	RRMSE	33.60	33.60	33.60	33.06	33.04	1.62	1.67
300	AB	20.90	20.90	20.90	20.60	20.59	1.43	1.46
	RRMSE	27.16	27.16	27.16	26.74	26.72	1.56	1.63

REFERENCES

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1996). Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis,* **28(4)**, 281-298.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1998). Geographically weighted regression - modelling spatial non-stationarity. *J. Roy. Statist. Soc.*, *Series D*, **47(3)**, 431- 443.

Chandra, H., Salvati, N., Chambers, R. and Tzavidis, N. (2012). Small area estimation under spatial nonstationarity. *Computational Statistics & Data Analysis*, **56(10)**, 2875-2888.

Chandra, H., Salvati, N., and Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. *Spatial Statistics*, **20**, 30-56.

Fotheringham, A.S., Charlton, M.E. and Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, **30(11)**, 1905-1927.

Leung, Y., Mei, C.L. and Zhang, W.X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, **32(1)**, 9-32.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377-387.

Valliant, R., Dorfman, A.H. and Royall, R.M. (2000). *Finite Population Sampling and Inference*. Wiley: New York.