stats

MDPI

Article

Selection of Auxiliary Variables for Three-Fold Linking Models in Small Area Estimation: A Simple and Effective Method

Song Cai * and J.N.K. Rao

School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6, Canada; jrao@math.carleton.ca
* Correspondence: song.cai@carleton.ca

Abstract: Model-based estimation of small area means can lead to reliable estimates when the area sample sizes are small. This is accomplished by borrowing strength across related areas using models linking area means to related covariates and random area effects. The effective selection of variables to be included in the linking model is important in small area estimation. The main purpose of this paper is to extend the earlier work on variable selection for area level and two-fold subarea level models to three-fold sub-subarea models linking sub-subarea means to related covariates and random effects at the area, sub-area, and sub-subarea levels. The proposed variable selection method transforms the sub-subarea means to reduce the linking model to a standard regression model and applies commonly used criteria for variable selection, such as AIC and BIC, to the reduced model. The resulting criteria depend on the unknown sub-subarea means, which are then estimated using the sample sub-subarea means. Then, the estimated selection criteria are used for variable selection. Simulation results on the performance of the proposed variable selection method relative to methods based on area level and two-fold subarea level models are also presented.

Keywords: Fay–Herriot model; information criterion; transformation; two-fold subarea model; variable selection

Citation: Cai, S.; Rao, J.N.K. Selection of Auxiliary Variables for Three-Fold Linking Models in Small Area Estimation: A Simple and Effective Method. *Stats* **2022**, *5*, 128–138. https://doi.org/10.3390/stats5010009

Academic Editor: Wei Zhu

Received: 9 January 2022
Accepted: 3 February 2022
Published: 5 February 2022

1. Introduction

Sample surveys are designed to provide reliable estimates of the overall means of a finite population and means for large domains or sub-populations (areas). For areas with small sample sizes (called small areas), direct area-specific estimators from the survey data are unreliable, and it is necessary to use model-based methods based on models linking area means to related covariates and random area effects. Resulting model-based estimators can lead to a significant increase in precision relative to direct estimators. Rao and Molina [1], in Chapter 6, provide a detailed account of model-based estimation under area level models. The effective selection of auxiliary variables to be included in the linking model is important for the success of model-based small area estimation (SAE).

A basic area level model, due to Fay and Herriot [2], is widely used for SAE in practice. Suppose that we have m areas with direct estimators y_i of the area means θ_i $(i = 1, \ldots, m)$ and associated candidate covariate vectors $\boldsymbol{x}_i$. The area level model consists of two components: a sampling model given by

$$y_i = \theta_i + e_i,\ i = 1, \ldots, m \tag{1}$$

and a linking model given by

$$\theta_i = \boldsymbol{x}_i^{\mathsf{T}} \boldsymbol{\beta} + u_i,\ i = 1, \ldots, m, \tag{2}$$

where e_i denotes sampling errors assumed to be independent $N(0, \Psi_i)$ with known sampling variance Ψ_i, and u_i denotes random area effects independent of e_i that are assumed to be independent and identically distributed (iid) as $N(0, \sigma_u^2)$ with unknown variance σ_u^2. In practice, the sampling variances are obtained by smoothing the estimators of sampling

variances using the Generalized Variance Function (GVF) method [3] and treating the smoothed estimators as the sampling variances Ψ_i. It is clear from (2) that it has a standard linear regression model form, and standard variable selection methods, such as Akaike Information Criterion (AIC) or Bayesian Information Criterions (BIC), can be applied to select variables, provided the area means θ_i are known. Lahiri and Suntornchost [4] estimated the resulting selection criteria using the sampling model (1) and proposed to use them for variable selection (see Section 2.1 for details). We refer the reader to Rao and Molina [1], Chapter 6, for details of empirical best linear unbiased prediction (EBLUP) estimators of area means from the models (1) and (2) for specified covariate vectors $\boldsymbol{x}_i$. The EBLUP estimator of θ_i is a weighted average of the direct estimator y_i and a synthetic regression estimator $\boldsymbol{x}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ denotes an estimator of the regression parameter vector $\boldsymbol{\beta}$. For a non-sampled area, direct estimator is not available. Hence, the synthetic estimator $\boldsymbol{x}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}}$ is used to estimate small area mean, provided the associated $\boldsymbol{x}_i$ is known. Fay and Herriot [2] obtained EBLUP estimates of per-capita income for small places in the USA, using the basic area level model given by (1) and (2).

Estimation of means for subareas nested within areas is of considerable interest. Mohadjer et al. [5] studied adult literacy for counties (subareas) sampled from states (areas), using data from the 2003 U.S. National Assessment of Adult Literacy. A two-fold subarea model is used to estimate subarea means θ_{ij} from n_i subareas j sampled from m areas i. A two-fold linking model on the subarea means θ_{ij} is given by

$$\theta_{ij} = \boldsymbol{x}_{ij}^{\mathsf{T}}\boldsymbol{\beta} + v_i + u_{ij},\ j = 1, \ldots, n_i;\ i = 1, \ldots, m, \tag{3}$$

where $\boldsymbol{x}_{ij}$ is the vector of covariates associated with θ_{ij}, and v_i is random area effect independent of random subarea effect u_{ij}. Furthermore, $v_i \overset{iid}{\sim} \mathrm{N}(0, \sigma_v^2)$ and $u_{ij} \overset{iid}{\sim} \mathrm{N}(0, \sigma_u^2)$. The linking model (3) is combined with the sampling model for the direct estimators y_{ij}, and it is given by

$$y_{ij} = \theta_{ij} + e_{ij},\ j = 1, \ldots, n_i;\ i = 1, \ldots, m, \tag{4}$$

where e_{ij} are sampling errors independently distributed as $\mathrm{N}(0, \Psi_{ij})$ with known sampling variances Ψ_{ij}, and assumed to be independent of v_i and u_{ij}. Torabi and Rao [6] obtained EBLUP estimators of subarea means for sampled subareas as well as non-sampled subareas. An advantage of the two-fold model is that the EBLUP estimator of a non-sampled subarea involves both the synthetic estimator of θ_{ij} and the direct estimators for the sampled subareas within the same area. For a non-sampled subarea within a non-sampled area, a synthetic estimator is used under the two-fold model. For variable selection under the two-fold model, Cai et al. [7] transformed the linking model to a standard regression model and applied variable selection criteria to the reduced model; see Section 2.2 for details.

Three-fold linking models involving sub-subareas (level 3) nested within subareas (level 2) which in turn are nested within areas (level 1) are also of practical interest. For example, such models were used in the Program for the International Assessment of Adult Competencies (PIAAC) in the context of estimating means for sub-subareas (counties) nested within subareas (states), which in turn are nested within areas (census divisions). Details of this application are reported in Krenzke et al. [8] and Ren et al. [9]. A three-fold linking model on the sub-subarea means θ_{ijk} is given by

$$\theta_{ijk} = \boldsymbol{x}_{ijk}^{\mathsf{T}}\boldsymbol{\beta} + w_i + v_{ij} + u_{ijk},\ k = 1, \ldots, n_{ij};\ j = 1, \ldots, m_i;\ i = 1, \ldots, L, \tag{5}$$

where k denotes sub-subarea nested within subarea j nested within area i, $\boldsymbol{x}_{ijk}$ is the vector of covariates associated with θ_{ijk}, w_i is the random area effect, v_{ij} is the random subarea effect, and u_{ijk} is the random sub-sub area effect. We assume that all the L areas in the population are included in the sample, but not all the subareas within an area are covered by the sample. Furthermore, not all the sub-subareas within a subarea covered by the sample are included in the sample. We assume that the three random effects in the model (5)

are independent, $w_i \overset{iid}{\sim} \mathrm{N}(0, \sigma_w^2)$, $v_{ij} \overset{iid}{\sim} \mathrm{N}(0, \sigma_v^2)$ and $u_{ijk} \overset{iid}{\sim} \mathrm{N}(0, \sigma_u^2)$. The linking model (5) is combined with the sampling model for the direct estimators y_{ijk} of the means θ_{ijk} for the sub-subareas in the sample. It is given by

$$y_{ijk} = \theta_{ijk} + e_{ijk},\ k = 1, \ldots, n_{ij};\ j = 1, \ldots, m_i;\ i = 1, \ldots, L, \tag{6}$$

where the e_{ijk} are sampling errors assumed to be independently distributed as $N(0, \Psi_{ijk})$ with known sampling variances Ψ_{ijk}, and they are assumed to be independent of the random effects w_i, v_{ij}, and u_{ijk}. In practice, the sampling variances are ascertained through smoothing of the estimated sampling variances, as done in the PIAAC project.

The survey design may not have the same hierarchical structure as the linking model (5). For example, in the PIAAC project, data from a stratified multistage sample with a different hierarchical structure are used. Given the vector of covariates $\boldsymbol{x}_{ijk}$ after variable selection, EBLUP estimators of the sub-subarea means can be obtained. It should be noted that the EBLUP estimators for non-sampled sub-subareas within a sampled subarea as well as those within non-sampled subareas avoid pure synthetic estimation by virtue of the area effects w_i included in the linking model (5), noting that all the areas in the population are included in the sample. In the PIAAC study, a hierarchical Bayes (HB) approach was used to estimate the population sub-subarea means. We will report EBLUP estimation for the three-fold model, which is given by (5) and (6), in a separate paper.

The main purpose of this paper is to extend the transformation method of Cai et al. [7] for variable selection to three-fold models given by (5) and (6). We propose two transformation-based methods—one is parameter free and the other is parameter-dependent—for variable selection. Section 2 is a review of some relevant variable selection methods for the area level model and the two-fold subarea model. Variable selection methods for the three-fold model are presented in Section 3. Results of a simulation study on the performance of the proposed methods relative to some naive alternatives, based on one-fold and two-fold models, are presented in Section 4. Some concluding remarks are presented in Section 5.

2. Area Level and Subarea Level Linking Models: Methods for Variable Selection

We now provide a brief review of earlier work on variable selection for area level and subarea level linking models related to the method for sub-subarea linking models presented in Section 3.

2.1. Area Level Model

The area level linking model (2) has the standard linear regression model form with unknown θ_i as the dependent variable. Lahiri and Suntornchost [4] noted that standard variable selection criteria applied to (2), such as AIC, BIC, and Mallow's C_p, are continuous functions of the unknown error mean sum of squares $\mathrm{MSE}_\theta = (m-p)^{-1}\theta^\intercal(I_m - P_X)\theta$, where $\theta = (\theta_1 \ \cdots \ \theta_m)^\intercal$, $P_X = X(X^\intercal X)^{-1}X^\intercal$ is the projection matrix with $X = (\boldsymbol{x}_1 \ \cdots \ \boldsymbol{x}_m)^\intercal$, I_m is the identity matrix of order m, and p is the dimension of $\boldsymbol{\beta}$. Then, the unknown MSE_θ is replaced by a consistent estimator obtained as

$$\widehat{\mathrm{MSE}}_\theta = \mathrm{MSE}_y - \overline{\Psi}_w, \tag{7}$$

where MSE_y is obtained by substituting $y = (y_1 \ \cdots \ y_m)^\intercal$ for θ in the above expression for MSE_θ and $\overline{\Psi}_w = (m-p)^{-1}\sum_{i=1}^m (1-h_{ii})\Psi_i$ is a weighted mean of the sampling variance Ψ_i with $h_{ii} = \boldsymbol{x}_i^\intercal (X^\intercal X)^{-1}\boldsymbol{x}_i$. The estimator (7) can take negative values, and modifications to (7) leading to positive values were proposed by Lahiri and Suntornchost [4].

As noted earlier, standard variable selection criteria applied to the linking model (2) are simple functions of MSE_θ and can be estimated by simply substituting MSE_y for MSE_θ. For example, BIC applied to linking model (2) can be estimated as

$$\widehat{\mathrm{BIC}} = m \log\left\{(m-p)\widehat{\mathrm{MSE}}_\theta / m\right\} + p \log m.$$

Some other variable selection criteria applicable to the area level model include the conditional AIC (cAIC) proposed by Han [10] and mixed generalized AIC (xGAIC) proposed by Lombardía et al. [11].

2.2. Subarea Level Model

Cai et al. [7] extended the method of Lahiri and Suntornchost [4] to the subarea model given by (3) and (4). In this case, the linking model (3) does not have a standard linear regression model form, because the error terms $v_i + u_{ij}$ within areas are correlated. As a result, we need to first transform the linking model (3) to a standard linear regression model form with iid errors. Specifically, we rewrite the subarea model in matrix form as $\boldsymbol{y}_i = \boldsymbol{\theta}_i + \boldsymbol{e}_i$ and $\boldsymbol{\theta}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\tau}_i$, $i = 1, \ldots, m$, where $\boldsymbol{y}_i = (y_{i1} \ \cdots \ y_{in_i})^\mathsf{T}$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1} \ \cdots \ \boldsymbol{x}_{in_i})^\mathsf{T}$, $\boldsymbol{\theta}_i = (\theta_{i1} \ \cdots \ \theta_{in_i})^\mathsf{T}$, $\boldsymbol{e}_i = (e_{i1} \ \cdots \ e_{in_i})^\mathsf{T}$, and $\boldsymbol{\tau}_i = v_i \mathbb{1}_{n_i} + \boldsymbol{u}_i$ with $\mathbb{1}_{n_i}$ being a vector of 1s of length n_i and $\boldsymbol{u}_i = (u_{i1} \ \cdots \ u_{in_i})^\mathsf{T}$. Cai et al. [7] proposed to find a matrix A_i for each $i = 1, \ldots, m$ such that the covariance matrix of $\boldsymbol{\tau}_i^* := A_i\boldsymbol{\tau}_i$ is a diagonal matrix with equal diagonal elements across $i = 1, \ldots, m$. Then, a transformed two-fold model is obtained:

$$\boldsymbol{y}_i^* = \boldsymbol{\theta}_i^* + \boldsymbol{e}_i^* \ \text{ and } \ \boldsymbol{\theta}_i^* = \boldsymbol{X}_i^*\boldsymbol{\beta} + \boldsymbol{\tau}_i^*, \ i = 1, \ldots, m, \tag{8}$$

where $\boldsymbol{y}_i^* = A_i\boldsymbol{y}_i$, $\boldsymbol{e}_i^* = A_i\boldsymbol{e}_i$, $\boldsymbol{\theta}_i^* = A_i\boldsymbol{\theta}_i$, and $\boldsymbol{X}_i^* = A_i\boldsymbol{X}_i$. Noting that (8) has the standard regression model form on the transformed variables $\boldsymbol{y}_i^*$ and $\boldsymbol{\theta}_i^*$, we can apply the method used for the area level model to obtain variable selection criteria.

Cai et al. [7] gave two methods for finding the transformation matrix A_i, one being parameter-free and the other relying on estimated model-parameter values. The parameter-free transformation follows the parameter-free transformation method proposed by Li and Lahiri [12] for selecting auxiliary variables under the unit-level nested error regression (NER) model. Observing that the covariance matrix of $\boldsymbol{\tau}_i^*$ is given by

$$\mathrm{Cov}(\boldsymbol{\tau}_i^*) = A_i\Sigma_i A_i^\mathsf{T} = \sigma_v^2 (A_i\mathbb{1}_{n_i})(A_i\mathbb{1}_{n_i})^\mathsf{T} + \sigma_u^2 A_i A_i^\mathsf{T},$$

one can choose an matrix A_i such that (a) $A_i\mathbb{1}_{n_i} = 0$, and (b) $A_iA_i^\mathsf{T}$ is a diagonal matrix whose diagonal elements are equal for all $i = 1, \ldots, m$. Cai et al. [7] proposed a numerical procedure to find the A_i matrix satisfying the above conditions. As a result of the linear constraint (a), the rank of A_i is $n_i - 1$ at most, and as a result, the transformed two-fold model loses one data point for each sampled area. The parameter-dependent method used by Cai et al. [7] is the well-known Fuller–Battese transformation [13]. In practice, the parameter-free transformation is more likely to be used because of its simplicity and not requiring the estimates of variance parameters.

3. Sub-Subarea Linking Models: Variable Selection

In this section, we present the proposed method for variable selection under the sub-subarea linking model. We extend the method of Cai et al. [7] for the two-fold model to the three-fold case.

We first express the sub-subarea linking and sampling models given by (5) and (6) as

$$\boldsymbol{\theta}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\eta}_i, \ i = 1, \ldots, L, \tag{9}$$

and

$$\boldsymbol{y}_i = \boldsymbol{\theta}_i + \boldsymbol{e}_i, \ i = 1, \ldots, L, \tag{10}$$

respectively, where $\boldsymbol{y}_i = (y_{ij1} \ y_{ij2} \ \cdots \ y_{im_in_{ij}})^\mathsf{T}$, $\boldsymbol{X}_i = (\boldsymbol{x}_{ij1} \ \boldsymbol{x}_{ij2} \ \cdots \ \boldsymbol{x}_{im_in_{ij}})^\mathsf{T}$, $\boldsymbol{\theta}_i = (\theta_{ij1} \ \theta_{ij2} \ \ldots \ \theta_{im_in_{ij}})^\mathsf{T}$, $\boldsymbol{e}_i = (e_{ij1} \ e_{ij2} \ \ldots \ e_{im_in_{ij}})^\mathsf{T}$ and

$$\boldsymbol{\eta}_i = w_i\mathbb{1}_{n_i} + \Omega_i\boldsymbol{v}_i + \boldsymbol{u}_i$$

with $n_i = \sum_{j=1}^{m_i} n_{ij}$, $\Omega_i = \text{diag}(\mathbb{1}_{n_{i1}}, \mathbb{1}_{n_{i2}}, \ldots, \mathbb{1}_{n_{im_i}})$, $\boldsymbol{v}_i = (v_{i1} \ \cdots \ v_{im_i})^\intercal$ and $\boldsymbol{u}_i = (u_{i11} \ u_{i12} \ \cdots \ u_{im_i n_{im_i}})^\intercal$. Note that $\boldsymbol{\eta}_i \sim \text{N}(0, \Sigma_i)$, where

$$\Sigma_i = \text{Cov}(\boldsymbol{\eta}_i) = \sigma_w^2 \mathbb{1}_{n_i} \mathbb{1}_{n_i}^\intercal + \sigma_v^2 \Omega_i \Omega_i^\intercal + \sigma_u^2 I_{n_i}. \tag{11}$$

As in the case of subarea linking model (3), the covariance matrix Σ_i of $\boldsymbol{\eta}_i$ in the linking model (9) does not have a diagonal structure. Following the idea of Cai et al. [7], we first transform the three-fold linking model (9) using a linear transformation so that the covariance matrix of the transformed $\boldsymbol{\eta}_i$ has a diagonal structure. For each area i, we obtain a matrix T_i that makes the transformed vector $\boldsymbol{\eta}_i^* := T_i \boldsymbol{\eta}_i$ have a diagonal covariance matrix with diagonal elements being a positive constant c for all $i = 1, \ldots, L$. Using the T_i, we transform the three-fold sampling model (10) and linking model (9) into

$$\boldsymbol{y}_i^* = \boldsymbol{\theta}_i^* + \boldsymbol{e}_i^*, \tag{12}$$

$$\boldsymbol{\theta}_i^* = \boldsymbol{X}_i^* \boldsymbol{\beta} + \boldsymbol{\eta}_i^*, \tag{13}$$

where $\boldsymbol{y}_i^* = T_i \boldsymbol{y}_i$, $\boldsymbol{\theta}_i^* = T_i \boldsymbol{\theta}_i$, $\boldsymbol{e}_i^* = T_i \boldsymbol{e}_i$ and $\boldsymbol{X}_i^* = T_i \boldsymbol{X}_i$. The transformed linking model (13) is a standard linear regression model with unknown dependent variable $\boldsymbol{\theta}_i^*$, and it shares the same $\boldsymbol{\beta}$ parameter with the original linking model (9). Then, we can use a bias-correction method similar to that of Lahiri and Suntornchost [4] to estimate an information criterion for (13) so as to select auxiliary variables. The details of the proposed transformation and bias-correction methods are given in the following subsections.

3.1. Transformation Methods

3.1.1. Parameter-Free Transformation

It is desirable that the transformation matrices T_i, $i = 1, \ldots, L$ do not rely on unknown parameter values. To find parameter-free T_i, we follow the idea used by Cai et al. [7] for the two-fold subarea model. By (11),

$$\text{Cov}(\boldsymbol{\eta}_i^*) = T_i \, \text{Cov}(\boldsymbol{\eta}_i) T_i^\intercal = \sigma_w^2 (T_i \mathbb{1}_{n_i})(T_i \mathbb{1}_{n_i})^\intercal + \sigma_v^2 (T_i \Omega_i)(T_i \Omega_i)^\intercal + \sigma_u^2 T_i T_i^\intercal.$$

If $T_i \mathbb{1}_{n_i} = 0$, $T_i \Omega_i = 0$ and $T_i T_i^\intercal$ is a diagonal matrix with equal diagonal elements, then $\text{Cov}(\boldsymbol{\eta}_i^*)$ will have the desired diagonal structure. Furthermore, $T_i \Omega_i = 0$ implies $T_i \mathbb{1}_{n_i} = 0$ because $\Omega_i \mathbb{1}_{m_i} = \mathbb{1}_{n_i}$. Therefore, it suffices to find T_i such that (i) $T_i \Omega_i = 0$, and (ii) $T_i T_i^\intercal$ is a diagonal matrix with equal diagonal elements across $i = 1, \ldots, L$. Since the above two conditions do not involve any parameter, a matrix T_i that satisfies them will be parameter free.

Recall that $\Omega_i = \text{diag}(\mathbb{1}_{n_{i1}}, \mathbb{1}_{n_{i2}}, \ldots, \mathbb{1}_{n_{im_i}})$, which is a matrix with full column rank m_i. Therefore, imposing $T_i \Omega_i = 0$ on T_i introduces m_i independent linear constraints on T_i, with one constraint for each sub-area j, $j = 1, \ldots, m_i$. To be specific, the constraint for subarea j is $T_i \omega_j = 0$, where ω_j is the jth, $j = 1, \ldots, m_i$, column of Ω_i. Consequently, the rank of T_i is at most $n_i - m_i$, and hence, each area i will lose m_i data points (or equivalently, each subarea will lose one data point) in the transformation. This is different from the parameter-free transformation for the two-fold subarea model discussed in Section 2.2, where each area loses a single data point in the transformation.

In the following, we provide a numerical algorithm to find T_i for area i, $i = 1, \ldots, L$, that satisfies the above requirements (i) and (ii).

Step 1: For each subarea j, $j = 1, \ldots, m_i$, of area i, fix a set of $n_{ij} - 1$ linearly independent vectors of length n_{ij}, denoted $b_{i1}, b_{i2}, \ldots, b_{i(n_{ij}-1)}$, that satisfies $b_{ik}^\intercal \mathbb{1}_{n_{ij}} = 0$ for $k = 1, \ldots, n_{ij} - 1$. For a given k, a valid choice for b_{ik} is the vector of length n_{ij} whose kth element is 1, last element is -1, and all other elements are 0. For example, if $n_{ij} = 5$, then we can take b_{ik}, $k = 1, 2, 3, 4$, as $b_{i1} = (1\,0\,0\,0\,-1)^\intercal$, $b_{i2} = (0\,1\,0\,0\,-1)^\intercal$, $b_{i3} = (0\,0\,1\,0\,-1)^\intercal$ and $b_{i4} = (0\,0\,0\,1\,-1)^\intercal$. Another possibility is to take b_{ik} to be the vector whose kth element is 1 and all the other elements equal $\frac{-1}{n_{ij}-1}$ if $n_{ij} > 1$.

Step 2: Apply the Gram–Schmidt process to $b_{i1}, \ldots, b_{i(n_{ij}-1)}$ to acquire a set of orthogonal vectors $a_{i1}, a_{i2}, \ldots, a_{i(n_{ij}-1)}$ with $a_{i1} = b_{i1}$ and $a_{ik} = b_{ik} - \sum_{l=1}^{k-1} \mathrm{Proj}_{a_{il}}(b_{ik})$ for $k = 2, \ldots, n_{ij} - 1$, where $\mathrm{Proj}_y(x) := \frac{x^\intercal y}{y^\intercal y} y$ is the projection of vector x onto the line spanned by vector y. Construct a $(n_{ij} - 1)$ by n_{ij} matrix, denoted T_{ij}, from $a_{i1}, \ldots, a_{i(n_{ij}-1)}$ as $T_{ij} = \left(\frac{a_{i1}}{\|a_{i1}\|} \cdots \frac{a_{i(n_{ij}-1)}}{\|a_{i(n_{ij}-1)}\|} \right)^\intercal$, where $\|\cdot\|$ is the Euclidean norm.

Step 3: Repeat Step 1 and Step 2 for all subareas $j = 1, \ldots, m_i$ of area i to obtain matrices $T_{i1}, \ldots, T_{im_i}$. Take $T_i = \mathrm{diag}(T_{i1}, \ldots, T_{im_i})$.

The T_i constructed using the above steps is parameter free. Step 1 generates a set of linear independent vectors b_{ik}, $k = 1, \ldots, m_i$, satisfying $b_{ik}^\intercal \mathbb{1}_{n_{ij}} = 0$. The Gram–Schmidt process in Step 2 produces a set of orthonormal vectors a_{ik}, $k = 1, \ldots, m_i$, based on b_{ik}, while carrying over the property $a_{ik}^\intercal \mathbb{1}_{n_{ij}} = 0$. Thus, the matrix T_{ij} constructed in Step 2 satisfies $T_{ij}\mathbb{1}_{n_{ij}} = 0$ and $T_{ij}T_{ij}^\intercal = I_{n_{ij}-1}$, which in turn guarantee that the matrix T_i defined in Step 3 satisfies the requirements (i) and (ii) with $T_i T_i^\intercal = I_{n_i - m_i}$. Under this transformation, we get $\mathrm{Cov}(\boldsymbol{\eta}_i^*) = \sigma_u^2 I_{n_i - m_i}$.

A parameter-free transformation matrix T_i that satisfies the constraints (i) and (ii) is not unique. However, we found that different choices of T_i yield similar results.

3.1.2. Parameter-Dependent Transformation

It is straightforward to obtain a transformation matrix T_i that depends on the model parameter values. Since $\mathrm{Cov}(\boldsymbol{\tau}_i^*) = T_i \Sigma_i T_i^\intercal$, where Σ_i is given by (11), we can take $T_i = c\Sigma_i^{-1/2}$, where $\Sigma_i^{-1/2}$ is the positive definite square-root matrix of Σ_i^{-1} and c is a non-zero constant. Then, $\mathrm{Cov}(\boldsymbol{\tau}_i^*)$ is a diagonal matrix whose diagonal elements are all equal to c. Note that Σ_i is determined by the variance parameters σ_w^2, σ_v^2 and σ_u^2, so this transformation matrix is parameter-dependent. Under the two-fold subarea model, applying this idea and choosing $c = \sigma_u$ yields the Fuller–Battese Transformation [7].

Under the three-fold sub-subarea model, we found that

$$\Sigma_i^{-1} = \sigma_u^{-2}\left(I_{n_i} - \Lambda_i - \xi_i \mathbb{1}_{n_i}\mathbb{1}_{n_i}^\intercal + \xi_i \mathbb{1}_{n_i}\mathbb{1}_{n_i}^\intercal \Lambda_i + \xi_i \Lambda_i \mathbb{1}_{n_i}\mathbb{1}_{n_i}^\intercal - \xi_i \Lambda_i \mathbb{1}_{n_i}\mathbb{1}_{n_i}^\intercal \Lambda_i\right),$$

where $\Lambda_i = \mathrm{diag}\left(\rho_{i1}\mathbb{1}_{n_{i1}}\mathbb{1}_{n_{i1}}^\intercal, \ldots, \rho_{im_i}\mathbb{1}_{n_{im_i}}\mathbb{1}_{n_{im_i}}^\intercal\right)$, $\rho_{ij} = \sigma_v^2/(\sigma_u^2 + n_{ij}\sigma_v^2)$ and $\xi_i = \sigma_w^2 / \left\{\sigma_u^2 + \sigma_w^2\left(n_i - \sum_{k=1}^{m_i} \rho_{ik} n_{ik}^2\right)\right\}$. The square-root matrix $\Sigma_i^{-1/2}$ has a complicated expression but can be found easily with a numerical procedure, for example, by applying the spectral decomposition or polar decomposition on Σ_i^{-1}. Taking $T_i = \sigma_u \Sigma_i^{-1/2}$, we get $\mathrm{Cov}(\boldsymbol{\tau}_i^*) = \sigma_u^2 I_{n_i}$.

In practice, we need to estimate the variance parameters σ_w^2, σ_v^2, and σ_u^2 to construct the transformation matrices T_i, as in the case of the subarea model given by (8).

3.2. Estimating Variable Selection Criteria: Sub-Subarea Model

After transformation, the linking model (13) takes the matrix form of a regular regression model with unobserved response variable values $\boldsymbol{\theta}_i^*$. We now use a method similar to that of Cai et al. [7] to estimate information criteria, including AIC, BIC, and Mallows' C_p,

for the transformed linking model (13). Then, these information criteria can be used for selecting auxiliary variables under the three-fold sub-subarea model.

The error mean sum of squares of the transformed linking model (13) is given by

$$\mathrm{MSE}_{\boldsymbol{\theta}^*} = \frac{1}{n^* - p}\boldsymbol{\theta}^{*\mathsf{T}}(I_{n^*} - P_{\boldsymbol{X}^*})\boldsymbol{\theta}^*,$$

where $\boldsymbol{\theta}^* = \left(\boldsymbol{\theta}_1^{*\mathsf{T}} \cdots \boldsymbol{\theta}_L^{*\mathsf{T}}\right)^{\mathsf{T}}$, $P_{\boldsymbol{X}^*} = \boldsymbol{X}^*\left(\boldsymbol{X}^{*\mathsf{T}}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*\mathsf{T}}$ with $\boldsymbol{X}^* = \left(\boldsymbol{X}_1^{*\mathsf{T}} \cdots \boldsymbol{X}_L^{*\mathsf{T}}\right)^{\mathsf{T}}$, n^* is the dimension of $\boldsymbol{\theta}^*$, and p is the dimension of $\boldsymbol{\beta}$. Since $\boldsymbol{\theta}^*$ are unobserved, $\mathrm{MSE}_{\boldsymbol{\theta}^*}$ cannot be calculated. Instead, we estimate $\mathrm{MSE}_{\boldsymbol{\theta}^*}$ based on the transformed sampling model (12). Let $\boldsymbol{y}^* = \left(\boldsymbol{y}_1^{*\mathsf{T}} \ \ldots \ \boldsymbol{y}_L^{*\mathsf{T}}\right)^{\mathsf{T}}$ and $\boldsymbol{e}^* = \left(\boldsymbol{e}_1^{*\mathsf{T}} \ \ldots \ \boldsymbol{e}_L^{*\mathsf{T}}\right)^{\mathsf{T}}$. Put

$$\mathrm{MSE}_{\boldsymbol{y}^*} = \frac{1}{n^* - p}\boldsymbol{y}^{*\mathsf{T}}(I_{n^*} - P_{\boldsymbol{X}^*})\boldsymbol{y}^*.$$

We propose to estimate $\mathrm{MSE}_{\boldsymbol{\theta}^*}$ by

$$\widehat{\mathrm{MSE}}_{\boldsymbol{\theta}^*} = \mathrm{MSE}_{\boldsymbol{y}^*} - \frac{1}{n^* - p}\,\mathrm{tr}\left\{(I_{n^*} - P_{\boldsymbol{X}^*})V_{\boldsymbol{e}^*}\right\}, \tag{14}$$

where $V_{\boldsymbol{e}^*}$ is the covariance matrix of $\boldsymbol{e}^*$, given by $V_{\boldsymbol{e}^*} = \mathrm{Cov}(\boldsymbol{e}^*) = TV_{\boldsymbol{e}}T^{\mathsf{T}}$ with $T = \mathrm{diag}(T_1, \ldots, T_L)$ and $V_{\boldsymbol{e}} = \mathrm{diag}(\Psi_{111}, \Psi_{112}, \ldots, \Psi_{Lm_Ln_{Lm_L}})$. It can be shown, using the same argument as used in the proof of Theorem 1 of Cai et al. [7], that if the sampling variances Ψ_{ijk} are bounded for all i, j and k, and $n_{ij} \geq 2$ for all i and j, then

$$\widehat{\mathrm{MSE}}_{\boldsymbol{\theta}^*} = \mathrm{MSE}_{\boldsymbol{\theta}^*} + o_p(1) \tag{15}$$

as the number of areas $L \to \infty$.

The term $(n^* - p)^{-1}\,\mathrm{tr}\left\{(I_{n^*} - P_{\boldsymbol{X}^*})V_{\boldsymbol{e}^*}\right\}$ in (14) can be considered as a bias-correction term, and because of its presence, $\widehat{\mathrm{MSE}}_{\boldsymbol{\theta}^*}$ may take a negative value. A simple truncation or a continuous transformation of $\widehat{\mathrm{MSE}}_{\theta^*}$ as suggested by Lahiri and Suntornchost [4] may be used to obtain a strictly positive estimate of MSE_{θ^*}.

Given the above estimator of MSE_{θ^*}, estimators of the AIC, BIC and Mallows' C_p for the transformed linking model (13) are readily constructed. The AIC, BIC and Mallows' C_p of a submodel of (13) with p_s covariates are given by

$$\mathrm{AIC}^{(s)} = n^* \log\left\{\left(n^* - p_s\right)\mathrm{MSE}_{\boldsymbol{\theta}^*}^{(s)}/n^*\right\} + 2p_s, \tag{16a}$$

$$\mathrm{BIC}^{(s)} = n^* \log\left\{\left(n^* - p_s\right)\mathrm{MSE}_{\boldsymbol{\theta}^*}^{(s)}/n^*\right\} + p_s\log(n^*), \tag{16b}$$

$$C_p^{(s)} = \left(n^* - p_s\right)\mathrm{MSE}_{\boldsymbol{\theta}^*}^{(s)} / \mathrm{MSE}_{\boldsymbol{\theta}^*} + 2p_s - n^*, \tag{16c}$$

respectively, where $\mathrm{MSE}_{\boldsymbol{\theta}^*}^{(s)}$ is the MSE from the submodel. Their estimators, denoted $\widehat{\mathrm{AIC}}^{(s)}$, $\widehat{\mathrm{BIC}}^{(s)}$ and $\widehat{C_p}^{(s)}$, respectively, are obtained by substituting $\widehat{\mathrm{MSE}}_{\boldsymbol{\theta}^*}$ into the corresponding expressions in (16a) to (16c). Then, variable selection is carried out by choosing one of the above criteria and estimating its values for a set of specified sub-models. The sub-model with the smallest estimated criterion value is chosen as the selected linking model. Noting that the criteria (16a)–(16c) are continuous functions of $\mathrm{MSE}_{\boldsymbol{\theta}^*}$ and the error of the estimator $\widehat{\mathrm{MSE}}_{\boldsymbol{\theta}^*}$ is of $o_p(1)$, it follows from the continuous mapping theorem [14] (Theorem 2.3) that the error in the estimated variable selection criteria is also $o_p(1)$ and hence negligible when the number of areas L is large.

4. Results of a Simulation Study

This section provides results of a limited simulation study on the performance of the proposed method for variable selection for sub-subarea linking models. The simulation

data are generated from the three-fold sub-subarea model given by (5) and (6). The number of areas is set to $L = 10$ and the number of subareas sampled from each area i, $i = 1, \ldots, L$, is set to $m_i = 5$. The number of sampled sub-subareas is taken as $n_{ij} = 8$ for every subarea j in areas $i = 1, \ldots, 5$, $n_{ij} = 5$ for every subarea in areas $i = 6, 7, 8$ and $n_{ij} = 10$ for each subarea in areas $i = 9, 10$. The sampling standard deviation $\sqrt{\Psi_{ijk}}$ in the sampling model (6) is generated from Unif$(0.5, 1.5)$. The standard deviation of the sub-subarea random effect in the linking model (5) is set to $\sigma_u = 2$. A few settings for the standard deviations of the area-level and subarea-level random effects, (σ_w, σ_v), are used: $(2, 2)$, $(4, 3)$, $(6, 3)$, $(8, 4)$, $(6, 6)$, $(3, 6)$ and $(4, 8)$. We consider a linking model that has an intercept term with corresponding covariate $x_{ijk,1} = 1$ and eight other covariates $x_{ijk,l}$ $(l = 2, \ldots, 9)$ generated as follows: $\log x_{ijk,2} \sim N(0.3, 0.5)$ with mean 0.3 and variance 0.5, $x_{ijk,3} \sim$ Gamma$(1.5, 2)$ with shape parameter 1.5 and rate parameter 2, $x_{ijk,4} \sim$ N$(0, 0.8)$, $x_{ijk,5} \sim$ N$(1, 1.5)$, $x_{ijk,6} \sim$ Gamma$(0.6, 10)$, $x_{ijk,7} \sim$ Beta$(0.5, 0.5)$ with shape parameters 0.5 and 0.5, $x_{ijk,8} \sim$ Unif$(1, 3)$ on the interval $(0, 3)$, and $x_{ij,9} \sim$ Poisson(1.5) with mean parameter 1.5. The value of the regression parameter vector $\boldsymbol{\beta}$ is set to $(2, 3, 0, 4, 0, 8, 0, 1, 0)^\intercal$. It corresponds to a true model consisting of the intercept term of value 2 and covariates $x_{ij,2}$, $x_{ij,4}$, $x_{ij,6}$ and $x_{ij,8}$. For variable selection, we always include the intercept term when we compare all possible sub-models defined by the inclusion/exclusion of the eight variables $x_{ij,2}, \ldots, x_{ij,9}$.

We generated 5000 simulation runs, and the covariates are generated first and kept fixed throughout all simulation runs. Then, we generated the response vectors $\boldsymbol{y}_i$, $j = 1, \ldots, L$, from the sub-subarea model given by (9) and (10) for each simulation run, using the specified settings.

We report the performance of the proposed method with parameter-free transformation ($3F_{pfree}$) and parameter-dependent transformation ($3F_{pdep}$). For $3F_{pdep}$, the true parameter values are used here, for simplicity. Under estimated parameter values, the performance of $3F_{pdep}$ is likely to be inferior. The parameter-free and parameter-dependent methods of Cai et al. [7] for the two-fold subarea model are used for comparison. To fit a two-fold subarea model to the data with a three-fold structure, the actual sub-subareas are treated as the subareas in the two-fold model. We can treat either (i) the actual subareas or (ii) the actual areas as the areas in the two-fold model. Treatment (i) is a natural choice when there is substantial subarea-level variability. Under treatment (i), where the actual subareas are treated as areas, the parameter-free transformation under the two-fold model is algebraically identical to the parameter-free transformation under the three-fold model. As a result, variable selection based on the parameter-free transformation under treatment (i) leads to the same set of variables as that under the three-fold model. However, it leads to pure synthetic estimates for non-sampled areas (actual subareas). Moreover, computationally, there is no advantage of treatment (i) over the three-fold model because the same transformation is used. On the other hand, the parameter-dependent method applied to treatment (i) may lead to a different set of variables. Therefore, we report the simulation results only for the parameter-dependent method under treatment (i), which is denoted as $2F\text{-}S\text{-}SS_{pdep}$. The two-fold parameter-free and parameter-dependent methods under treatment (ii) are denoted as $2F\text{-}A\text{-}SS_{pfree}$ and $2F\text{-}A\text{-}SS_{pdep}$, respectively. Under treatment (ii), pure synthetic estimation is avoided because all areas are sampled. For comparison, we further consider three naive methods designed for the one-fold FH model and the regular linear regression model, including the Lahiri–Suntornchost [4] method (Naive-LS) and Han's [10] cAIC method (Naive-cAIC) for the FH model, as well as an information criterion-based method for the regular linear regression model fitted naively to the data (Naive-LM). For Naive-LS and Naive-cAIC, the actual sub-subareas are treated as the areas in the FH model. For Naive-LM, the sub-subarea level direct estimator y_{ijk} is treated as the response variable of the regular linear regression model.

Table 1 summarizes the simulation results for variable selection using BIC.

Table 1. True model selection rate (%): BIC.

Method	(σ_w, σ_v)						
	(2, 2)	**(4, 3)**	**(6, 3)**	**(8, 4)**	**(6, 6)**	**(3, 6)**	**(4, 8)**
$3F_{pfree}$	87.12	87.62	87.50	88.18	87.26	87.32	87.02
$3F_{pdep}$	87.94	88.20	88.46	88.52	88.00	87.96	87.60
$2F\text{-}S\text{-}SS_{pdep}$	87.64	87.82	88.16	88.48	87.90	87.86	87.56
$2F\text{-}A\text{-}SS_{pfree}$	83.28	63.14	62.62	36.70	8.38	9.22	2.24
$2F\text{-}A\text{-}SS_{pdep}$	82.60	60.84	60.66	34.58	7.24	8.48	1.80
Naive-LS	63.62	19.68	8.80	2.56	1.94	4.96	0.78
Naive-LM	60.94	18.32	8.26	2.44	1.84	4.70	0.76

The proposed $3F_{pfree}$ and $3F_{pdep}$ perform equally well with a stable rate between 87% and 89% in selecting the true model under all settings for (σ_w, σ_v). The two-fold method $2F\text{-}S\text{-}SS_{pdep}$, which treats the actual subareas as areas in the two-fold model, exhibits similar performance to that of the proposed methods. All the other methods have inferior performance and display a dramatic decay in rate of selecting the true model when σ_w and σ_u increase. This indicates that in the presence of strong area-level effect or subarea-level effect, which often happens in practice, $3F_{pfree}$, $3F_{pdep}$ and $2F\text{-}S\text{-}SS_{pdep}$ are preferred over the other alternative methods.

The simulation results based on AIC and Naive cAIC are given in Table 2.

Table 2. True model selection rate (%): AIC and Naive-cAIC.

Method	(σ_w, σ_v)						
	(2, 2)	**(4, 3)**	**(6, 3)**	**(8, 4)**	**(6, 6)**	**(3, 6)**	**(4, 8)**
$3F_{pfree}$	43.88	42.84	43.76	43.92	43.38	44.02	43.42
$3F_{pdep}$	44.22	43.18	43.66	44.08	43.36	44.04	43.48
$2F\text{-}S\text{-}SS_{pdep}$	44.32	43.14	43.66	43.68	43.22	44.02	43.32
$2F\text{-}A\text{-}SS_{pfree}$	47.00	46.60	47.10	42.86	26.46	27.16	14.60
$2F\text{-}A\text{-}SS_{pdep}$	47.78	48.48	48.36	43.96	26.36	26.52	14.54
Naive-LS	45.00	31.74	22.48	14.62	14.02	19.84	10.52
Naive-LM	47.02	32.18	22.54	14.56	13.94	19.76	10.40
Naive-cAIC	42.86	26.64	17.28	12.12	11.16	15.62	8.84

Compared with BIC, AIC gives a significantly lower true-model selection rate under all the methods. As the case for BIC, methods $3F_{pfree}$, $3F_{pdep}$ and $2F\text{-}S\text{-}SS_{pdep}$ perform equally well and yield stable results for different (σ_w, σ_v) values, and they have better performance than the other methods. Methods $2F\text{-}A\text{-}SS_{pfree}$ and $2F\text{-}A\text{-}SS_{pdep}$ have slightly better performance than $3F_{pfree}$, $3F_{pdep}$ and $2F\text{-}S\text{-}SS_{pdep}$ when $(\sigma_w, \sigma_v) = (2, 2), (4, 3)$, and $(6, 3)$ but notably inferior performance under the other settings for (σ_w, σ_v). Methods Naive-LS, Naive-LM and Naive-cAIC have significantly lower rates of selecting the true model than the other methods.

Table 3 reports simulation results under Mallows' C_p criterion for variable selection. The results in Table 3 are similar to those reported in Table 2 under AIC, and the same conclusions hold.

Table 3. True model selection rate (%): Mallows' C_p.

Method	(σ_w, σ_v)						
	(2, 2)	**(4, 3)**	**(6, 3)**	**(8, 4)**	**(6, 6)**	**(3, 6)**	**(4, 8)**
$3F_{pfree}$	44.78	43.66	44.84	44.60	44.00	44.84	44.04
$3F_{pdep}$	45.02	43.90	44.40	44.90	44.16	44.80	44.30
2F-S-SS_{pdep}	44.96	43.74	44.32	44.50	43.76	44.74	44.10
2F-A-SS_{pfree}	47.60	47.28	47.84	43.32	26.34	27.08	14.54
2F-A-SS_{pdep}	48.82	49.08	49.38	44.52	26.54	26.64	14.22
Naive-LS	45.60	32.16	22.60	14.52	13.98	19.80	10.32
Naive-LM	47.66	32.52	22.64	14.54	14.04	19.80	10.24

5. Concluding Remarks

A transformation-based method is proposed for selecting covariates under the three-fold sub-subarea model for small area estimation. Two transformations, one being parameter-free and the other being parameter-dependent, are proposed to accompany the variable selection method. Compared to the parameter-free transformation, the parameter-dependent transformation does not induce loss of data points but requires estimated variance parameters in practice. We prefer the parameter-free transformation for its simplicity and not requiring the estimates of variance parameters. The performance of parameter-free and parameter-dependent transformation methods is similar under various simulation settings for variances of the area-level and subarea-level random effects. EBLUP estimation of sub-subarea means for sampled sub-subareas, non-sampled sub-subareas within sampled subareas, and sub-subareas within non-sampled subareas will be studied in detail in a separate paper. Measures of uncertainty of the EBLUP estimators will also be studied.

Author Contributions: Formal analysis, S.C.; Methodology, S.C., J.N.K.R.; Writing-original draft, S.C.; Writing-review & editing, S.C., J.N.K.R.; Conceptualization, J.N.K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by research grants to Song Cai and J.N.K. Rao from the Natural Sciences and Engineering Research Council of Canada.

Acknowledgments: We thank the reviewers for their useful comments and constructive suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rao, J.N.K.; Molina, I. *Small Area Estimation*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2015.
2. Fay, R.E.; Herriot, R.A. Estimates of income for small places: An application of james-stein procedures to census data. *J. Am. Stat. Assoc.* **1979**, *74*, 269–277. [CrossRef]
3. Wolter, K.M. *Introduction to Variance Estimation*, 2nd ed.; Springer: New York, NY, USA, 2007.
4. Lahiri, P.; Suntornchost, J. Variable selection for linear mixed models with applications in small area estimation. *Sankhyā B* **2015**, *77*, 312–320. [CrossRef]
5. Mohadjer, L.; Rao, J.N.K.; Liu, B.; Krenzke, T.; Van de Kerckhove, W. Hierarchical Bayes small area estimates of adult literacy using unmatched sampling and linking models. *J. Indian Soc. Agric.* **2012**, *66*, 55–63.
6. Torabi, M.; Rao, J.N.K. On small area estimation under a sub-area level model. *J. Multivar. Anal.* **2014**, *127*, 36–55. [CrossRef]
7. Cai, S.; Rao, J.N.K.; Dumitrescu, L.; Chatrchi, G. Effective transformation-based variable selection under two-fold subarea models in small area estimation. *Stat. Transit. New Ser.* **2020**, *21*, 68–83. [CrossRef]
8. Krenzke, T.; Mohadjer, L.; Li, J.; Erciulescu, A.; Fay, R.E.; Ren, W.; VanDeKerckhove, W.; Li, L.; Rao, J.N.K. *Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report*; Technical Report; Institute of Education Sciences, National Center for Education Statistics: Washington, DC, USA, 2020.
9. Ren, W.; Li, J.; Erciulescu, A.; Krenzke, T.; Mohadjer, L. A variable selection method for small area estimation modeling of the proficiency of adult competency. In Proceedings of the Survey Research Methods Section, Joint Statistical Meetings of the American Statistical Association, Alexandria, VA, USA, 2–6 August 2020; pp. 924–956.
10. Han, B. Conditional Akaike information criterion in the Fay-Herriot model. *Stat. Methodol.* **2013**, *11*, 53–67. [CrossRef]

11. Lombardía, M.J.; López-Vizcaíno, E.; Rueda, C. Mixed generalized akaike information criterion for small area models. *J. R. Stat. Soc. Ser. A* **2017**, *180*, 1229–1252. [CrossRef]
12. Li, Y.; Lahiri, P. A simple adaptation of variable selection software for regression models to select variables in nested error regression models. *Sankhyā B* **2019**, *81*, 302–317. [CrossRef]
13. Fuller, W.A.; Battese, G.E. Transformations for estimation of linear models with nested-error structure. *J. Am. Stat. Assoc.* **1973**, *68*, 626–632. [CrossRef]
14. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, MA, USA, 1998.