



***Centre for Statistical and Survey Methodology***

**The University of Wollongong**

**Working Paper**

01-11

A Semiparametric Block Bootstrap for Clustered Data

Ray Chambers and Hukum Chandra

*Copyright © 2008 by the Centre for Statistical & Survey Methodology, UOW. Work in progress, no part of this paper may be reproduced without permission from the Centre.*

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: [anica@uow.edu.au](mailto:anica@uow.edu.au)

# A Semiparametric Block Bootstrap for Clustered Data

Ray Chambers and Hukum Chandra

Centre for Statistical and Survey Methodology, University of Wollongong,  
Wollongong, NSW 2522, Australia.

E-mail: [ray@uow.edu.au](mailto:ray@uow.edu.au) and [hchandra@uow.edu.au](mailto:hchandra@uow.edu.au)

## ABSTRACT

Random effects models for hierarchically dependent data, e.g. clustered data, are widely used. A popular bootstrap method for such data is the parametric bootstrap based on the same random effects model as that used in inference. However, it is hard to justify this type of bootstrap when this model is known to be an approximation. In this paper we describe a semiparametric block bootstrap approach for clustered data that is simple to implement, free of both the distribution and the dependence assumptions of the parametric bootstrap and is consistent when the mixed model assumptions are valid. Results based on Monte Carlo simulation show that the proposed method seems robust to failure of the dependence assumptions of the assumed mixed model. An application to a realistic environmental data set indicates that the method produces sensible results.

**Key Words:** Hierarchical data; Correlated clusters; Block bootstrap; Confidence interval; Consistency; Nonparametric bootstrap.

## 1. INTRODUCTION

The bootstrap technique (Efron 1979; Efron and Tibshirani 1993) was originally developed for parametric inference given independent and identically distributed (*iid*) data. However, random effects models for hierarchically dependent data, e.g. clustered or multilevel data, are now in wide use. With such data, it is important to use bootstrap techniques that replicate the hierarchical dependence structure of the data. A popular way of achieving this is to use a parametric bootstrap based on the assumed hierarchical random effects model. This is usually very effective provided this model is correctly specified. On the other hand, if the stochastic assumptions of the model, e.g. the assumption that the random effects are *iid* Gaussian random variables, are violated, then it is hard to justify use of the parametric bootstrap. See for example, Rasbash et al. (2000). This is of particular concern since the bootstrap is often recommended as an alternative approach that is likely to lead to confidence intervals with better coverage in situations where the distribution assumptions that underly analytical methods are questionable (Davison and Hinkley 1997).

Much of the early research on bootstrapping clustered data was within the design-based framework for sample survey inference, where the main focus is on replicating the impact of various forms of cluster sampling on repeated sampling inference for finite population parameters. See Rao and Wu (1988) and Canty and Davison (1999). However, our approach in this paper is model-based, in the sense that we treat the clusters as part of the data generation mechanism rather the sampling scheme, and so is similar to the approach set out in Field and Welsh (2007). In particular, we consider inference with respect to the population model rather the sampling mechanism, and so our focus is on bootstrap inference for model parameters that accommodates the hierarchical dependence structure in the data. In this context, Carpenter et al. (2003) (hereafter CGR) describe a two level bootstrap for a random

effects model, while Field and Welsh (2007) review various approaches to bootstrapping clustered data.

In what follows we propose a semiparametric block bootstrap method for clustered, hierarchical or multilevel data. Our approach is semiparametric, in the sense that the marginal model is generated parametrically within the bootstrap while the dependence structure of the model residuals is generated nonparametrically. Furthermore, the proposed bootstrap is simple to implement and seems free of both the distribution and the dependence assumptions of the parametric bootstrap, with its main assumption being that the marginal model is correctly specified. Note that the block bootstrap itself is not new, since this method has been used extensively with spatial and time series data. See Clark and Allingham (2011), Hutson (2004), and Hall et al. (1995). However, to the best of our knowledge, there have been no previous applications of the block bootstrap idea to multilevel data.

The remainder of this paper is structured as follows. In Section 2 we describe how the parametric bootstrap and the CGR bootstrap can be used to construct bootstrap confidence intervals for multilevel data. We then describe a semiparametric block bootstrap method for such data and prove the consistency of the bootstrap confidence intervals obtained under this approach. Empirical results from model-based simulations of these different bootstrap methods are described in Section 3. In Section 4 we present results from the application of these bootstrap methods to a realistic environmental data set where the hierarchical model is at best an approximation. Finally, Section 5 concludes the paper with a summary of our major findings and a discussion of avenues for future research.

## **2. BOOTSTRAP METHODS FOR MULTILEVEL DATA**

In this Section we outline bootstrap methods for constructing confidence intervals for parameters of multilevel models, focusing on the two-level case. To this end, consider the

situation where we have data on a variable of interest  $y$  and a set of covariates  $\mathbf{x}$  for  $n$  individuals clustered within  $D$  groups. Following standard practice, we refer to individuals as level 1 units and clusters as level 2 units. There are  $n_i$  ( $i=1, \dots, D$ ) level 1 units making up level 2 unit  $i$  in the sample, with overall sample size  $n = \sum_{i=1}^D n_i$ . Such hierarchically structured data are commonly modelled using random effects. In this paper we focus on a linear random intercepts model of form

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij}, \quad j = 1, \dots, n_i; i = 1, \dots, D, \quad (1)$$

where  $y_{ij}$  denote the value of  $y$  for unit  $j$  in group  $i$ ,  $\mathbf{x}_{ij}$  is a  $p \times 1$  vector of auxiliary variables for unit  $j$  in group  $i$ ,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of regression coefficients,  $u_i$  denotes a cluster-specific (level 2) random effect and  $e_{ij}$  is an individual (level 1) random effect. We assume that  $\mathbf{x}_{ij}$  contains an intercept term as its first component. It is standard practice to model the random effects as Gaussian, and so we further assume that these effects are mutually independent between individuals and between clusters, with  $u_i \sim N(0, \sigma_u^2)$  and  $e_{ij} \sim N(0, \sigma_e^2)$ . It follows that  $E(y_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta}$  and  $Cov(y_{ij}, y_{ik}) = \sigma_u^2 + \sigma_e^2 I(j=k)$ , where  $I(A)$  is the indicator function for the event  $A$ . Let  $\mathbf{y}$  denote the  $n \times 1$  vector of values  $y_{ij}$  with  $\mathbf{x}$  denoting the corresponding  $n \times p$  matrix defined by the  $\mathbf{x}_{ij}$ . Then  $E(\mathbf{y}) = \mathbf{x}\boldsymbol{\beta}$  and  $\mathbf{V} = Var(\mathbf{y}) = diag\{V_i = \sigma_u^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T + \sigma_e^2 \mathbf{I}_{n_i}; i = 1, \dots, D\}$ , where  $\mathbf{I}_t$  and  $\mathbf{1}_t$  denote the identity matrix of order  $t$  and a  $t \times 1$  vector of ones respectively. The parameters  $\delta = (\sigma_u^2, \sigma_e^2)$  are typically referred to as the variance components of (1). Standard methods such as maximum likelihood (ML) or restricted maximum likelihood (REML) are used for estimating the unknown parameters of (1); see Harville (1977). In what follows we use a ‘hat’ to denote these estimates, i.e. we let  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$  denote the ML or REML estimates of

$\theta = (\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2)$ . These allow us to define empirical best linear unbiased predictors (EBLUPs)  $\hat{u}_i$  for the area effects  $u_i$ .

Given a bootstrap distribution for a component of  $\hat{\theta}$ , there are a number of methods that can be used to construct corresponding bootstrap confidence intervals for the parameter in  $\theta$  corresponding to that component. For reviews of these methods, see Efron and Tibshirani (1993), DiCiccio and Efron (1996), Davison and Hinkley (1997) and DiCiccio and Romano (1988). Here we use the percentile method, where a  $100(1-\alpha)$  percent bootstrap confidence interval for any component of  $\theta$  is constructed as the interval between upper and lower  $\alpha/2$  percentile values of the bootstrap distribution of that component. Taking some liberties with notation, let  $\hat{\theta}_{L,\alpha/2}$  denotes the bootstrap estimate for a parameter in  $\theta$  such that a fraction  $\alpha/2$  of all its bootstrap estimates are smaller than  $\hat{\theta}_{L,\alpha/2}$ , with  $\hat{\theta}_{U,\alpha/2}$  denoting the bootstrap estimate such that a fraction  $\alpha/2$  of all its bootstrap estimates are larger than  $\hat{\theta}_{U,\alpha/2}$ . Then an approximate  $100(1-\alpha)$  percent confidence interval for this parameter is  $[\hat{\theta}_{L,\alpha/2}, \hat{\theta}_{U,\alpha/2}]$ .

## 2.1 PARAMETRIC 2-LEVEL BOOTSTRAP

The parametric bootstrap method for the ML/REML estimates  $\hat{\theta} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$  obtained by fitting the model (1) to data with 2-level structure is defined as follows.

1. Generate independent level 2 errors for the  $D$  groups as  $u_i^* \sim N(0, \hat{\sigma}_u^2)$ ,  $i = 1, \dots, D$  and generate independent level 1 errors for all  $n$  sampled units as  $e_{ij}^* \sim N(0, \hat{\sigma}_e^2)$ ,  $j = 1, \dots, n_i; i = 1, \dots, D$ .
2. Simulate bootstrap sample data  $(y_{ij}^*, x_{ij})$  using the model  $y_{ij}^* = x_{ij}^T \hat{\boldsymbol{\beta}} + u_i^* + e_{ij}^*$ .

3. Fit the two level random effects model (1) to the bootstrap sample data generated in step 2 to obtain bootstrap parameter estimates  $\hat{\theta}^* = (\hat{\boldsymbol{\beta}}^*, \hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})$ .
4. Repeat steps 1-3  $B$  times to obtain  $B$  sets of bootstrap parameter estimates.

As noted in Section 1, this method works very well provided the model (1) holds. However, it is hard to justify this type of bootstrap if the stochastic assumptions of this model, e.g. that the random effects are *iid* Gaussian random variables, are violated.

## 2.2 SEMIPARAMETRIC 2-LEVEL BOOTSTRAP (CGR)

Carpenter et al. (2003) describe a bootstrap method for multilevel data that is less sensitive to model assumptions than the parametric bootstrap. As usual, we suppose that we have estimates  $\hat{\theta} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$  of  $\theta = (\boldsymbol{\beta}, \sigma_u^2, \sigma_e^2)$ . Note that this means that we also have the corresponding EBLUPs  $\hat{u}_i$  of  $u_i$  ( $i = 1, \dots, D$ ). In what follows we use the notation  $srswr(A, m)$  to indicate the outcome of taking a simple random sample of size  $m$  with replacement from the set  $A$ . The CGR bootstrap is then implemented as follows:

1. The  $D$  EBLUPs  $\hat{u}_i$  of the random effects  $u_i$  and the corresponding  $n$  level 1 residuals

$\hat{e}_{ij} = y_{ij} - x_{ij}^T \hat{\boldsymbol{\beta}} - \hat{u}_i$  are first scaled to ensure that they have variances equal to  $\hat{\sigma}_u^2$  and

$\hat{\sigma}_e^2$  respectively. The scaled level 2 residuals are  $\hat{u}_i^c = \hat{\sigma}_u \left\{ D^{-1} \left( \sum_i \hat{u}_i^2 \right) \right\}^{-1/2} \hat{u}_i$  and the

scaled level 1 residuals are  $\hat{e}_{ij}^c = \hat{\sigma}_e \left\{ n^{-1} \left( \sum_i \hat{e}_{ij}^2 \right) \right\}^{-1/2} \hat{e}_{ij}$ . Both sets of scaled residuals

are then centred at zero.

2. Sample independently with replacement from  $\hat{\mathbf{u}}^c = (\hat{u}_i^c)$  and  $\hat{\mathbf{e}}^c = (\hat{e}_{ij}^c)$  to get bootstrap

samples  $\mathbf{u}^*$  and  $\mathbf{e}^*$  of  $D$  level 2 residuals and  $n$  level 1 residuals respectively. That is,

$$\mathbf{u}^* = (u_i^*) = srswr\{\hat{\mathbf{u}}^c, D\} \text{ and } \mathbf{e}^* = (e_{ij}^*) = srswr\{\hat{\mathbf{e}}^c, n\}.$$

3. Simulate bootstrap sample data  $(y_{ij}^*, x_{ij})$  using the model  $y_{ij}^* = x_{ij}^T \hat{\boldsymbol{\beta}} + u_i^* + e_{ij}^*$ .
4. Fit the two level random effects model (1) to the bootstrap sample data generated in step 3 to obtain bootstrap parameter estimates  $\hat{\theta}^* = (\hat{\boldsymbol{\beta}}^*, \hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})$ .
5. Repeat steps 2-4  $B$  times to obtain  $B$  sets of bootstrap parameter estimates.

## 2.3 SEMIPARAMETRIC 2-LEVEL BLOCK BOOTSTRAP

Although the bootstrap errors used in the CGR approach are less sensitive to the stochastic assumptions of (1), they still rely on the model-based EBLUPs  $\hat{u}_i$  of the level 2 random effects. In addition, both the parametric and the CGR approaches assume homogeneity of within cluster variability. In practice, within cluster errors may not be homogeneous. For example, these errors can be correlated in environmental and agricultural applications, reflecting unmeasured spatial variation. Provided the within block residual heterogeneity is similar from cluster to cluster, we can use a block bootstrap approach to recreate this heterogeneity in our bootstrap. We therefore now describe a semiparametric block bootstrap approach that allows for such residual heterogeneity. This approach is semiparametric in the sense that although the marginal bootstrap model is based on the parametric fit to the sample data, the dependence structure in the model residuals is generated nonparametrically.

### 2.3.1 Semiparametric block bootstrap (SBB)

We first describe a simple semiparametric block (SBB) bootstrap for two-level data and then develop refinements to this method. The steps in the SBB bootstrap are as follows.

1. Using the marginal residuals:  $r_{ij} = y_{ij} - x_{ij}^T \hat{\boldsymbol{\beta}}$ ,  $j = 1, \dots, n_i; i = 1, \dots, D$ , calculate the level 2 average residuals for each of the  $D$  groups:  $\bar{r}_h = n_h^{-1} \sum_{j=1}^{n_h} r_{hj}$ ,  $h = 1, \dots, D$  and the level 1 residuals within each group  $h$  as  $r_{hj}^{(1)} = r_{hj} - \bar{r}_h$ ,  $j = 1, \dots, n_h; h = 1, \dots, D$ . Let  $\bar{\mathbf{r}}^{(2)}$  and  $\mathbf{r}_h^{(1)}$



denote the vector of  $D$  level 2 average residuals  $\bar{r}_h$  and the vector of  $n_h$  level 1 residuals  $r_{hj}^{(1)}$  for group  $h$  respectively.

2. Sample independently and with replacement from these two sets of residuals in order to define bootstrap errors for levels 1 and 2. In particular, level 2 bootstrap errors are given by  $\mathbf{r}^{*(2)} = (r_i^{*(2)}) = srswr(\bar{\mathbf{r}}^{(2)}, D)$ , while level 1 bootstrap errors in cluster  $i$  are given by  $\mathbf{r}_i^{*(1)} = (r_{ij}^{*(1)}) = srswr(\mathbf{r}_{h(i)}^{(1)}, n_i)$ , where  $h(i) = srswr(\{1, \dots, D\}, 1)$ .
3. Simulate bootstrap sample data  $(y_{ij}^*, x_{ij})$  using the model  $y_{ij}^* = x_{ij}^T \hat{\boldsymbol{\beta}} + r_i^{*(2)} + r_{ij}^{*(1)}$ .
4. Fit the two level random effects model (1) to the bootstrap sample data generated in step 3 to obtain bootstrap parameter estimates  $\hat{\boldsymbol{\theta}}^* = (\hat{\boldsymbol{\beta}}^*, \hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})$ .
5. Repeat steps 2-4  $B$  times to obtain  $B$  sets of bootstrap parameter estimates.

### 2.3.2 Use of centred and scaled residuals before bootstrapping (SBB.Prior)

In the semiparametric block bootstrap SBB described in the previous subsection, we note that  $E^*(r_i^{*(2)}) \neq 0$  and  $Var^*(r_i^{*(2)}) \neq \hat{\sigma}_u^2$ , where  $E^*$ ,  $Var^*$  denote expectation and variance respectively with respect to the bootstrap distribution generated under SBB. Consequently  $E^*(y_{ij}^*) \neq x_{ij}^T \hat{\boldsymbol{\beta}}$ , implying that the bootstrap confidence intervals generated by SBB are not consistent. An alternative, which also satisfies the conditions for consistency (Shao and Tu, 1995, Chapter 4), is to zero centre and scale residuals prior to their use in the bootstrap. That is, following the same procedure as used in the CGR bootstrap, the residuals  $\bar{r}_h$  and  $r_{hj}^{(1)}$  computed in step 1 of the SBB are transformed to zero-centred and scaled values  $\bar{r}_h^c = \hat{\sigma}_u \left\{ D^{-1} \left( \sum_i \bar{r}_i^2 \right) \right\}^{-1/2} \bar{r}_h$  and  $r_{hj}^{(1)c} = \hat{\sigma}_e \left\{ n^{-1} \left( \sum_i (r_{ij}^{(1)})^2 \right) \right\}^{-1/2} r_{hj}^{(1)}$  respectively before initiating the bootstrap process of steps 2-5, in which case we have  $E^*(r_i^{*(2)}) = E^*(r_{ij}^{*(1)}) = 0$ ,

$Var^*(r_i^{*(2)}) = \hat{\sigma}_u^2$  and  $Var^*(r_{ij}^{*(1)}) = \hat{\sigma}_e^2$ . Zero centring and scaling residuals prior to initiating the bootstrap ensures that the confidence intervals generated by the SBB are consistent. Because these residual adjustments are carried out before the bootstrap process, we refer to this method as SBB.Prior in what follows.

### 2.3.3 Tilting and tethering adjustments after bootstrapping (SBB.Post)

The variance components estimates  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_e^2$  should be asymptotically uncorrelated. However, there is no guarantee that the bootstrap estimates of these parameters generated by SBB are empirically uncorrelated. Furthermore, although SBB has the property of preserving residual within cluster heterogeneity, there is no guarantee that it preserves the observed between and within cluster variances. Both of these properties can be guaranteed by appropriately modifying the bootstrap distributions generated by SBB. We therefore now describe two further steps in the SBB procedure that ensure these properties.

- We first modify the bootstrap distributions of the logarithms of the variance components estimates so that they are empirically uncorrelated. The steps in this process are as follows:

- Let  $(\log \hat{\sigma}_u^{2*})$  and  $(\log \hat{\sigma}_e^{2*})$  denote the  $B$  vectors of bootstrap values of  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_e^2$  respectively. Define the  $B \times 2$  matrices

$$\mathbf{S}^* = \left[ (\log \hat{\sigma}_u^{2*}), (\log \hat{\sigma}_e^{2*}) \right]$$

$$\mathbf{M}^* = \left[ av(\log \hat{\sigma}_u^{*2}) \times \mathbf{1}_B, av(\log \hat{\sigma}_e^{*2}) \times \mathbf{1}_B \right]$$

$$\text{and } \mathbf{D}^* = \left[ sd(\log \hat{\sigma}_u^{*2}) \times \mathbf{1}_B, sd(\log \hat{\sigma}_e^{*2}) \times \mathbf{1}_B \right].$$

Here  $avS$  and  $sdS$  denote the average and standard deviation of the values in the vector  $S$ ,  $\mathbf{1}_B$  denotes a  $B$  vector of ones and  $\times$  denotes component-wise multiplication.

ii. Calculate the  $2 \times 2$  covariance matrix  $\mathbf{C}^* = \text{cov}(\mathbf{S}^*)$  and put

$$\mathbf{L}^* = \mathbf{M}^* + \left\{ (\mathbf{S}^* - \mathbf{M}^*) \mathbf{C}^{*-1/2} \right\} \times \mathbf{D}^* .$$

iii. The modified bootstrap values of  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_e^2$  (denoted  $\hat{\sigma}_u^{*\text{mod}2}$  and  $\hat{\sigma}_e^{*\text{mod}2}$  below) are then obtained by exponentiating the elements of  $\mathbf{L}^*$ .

- All bootstrap distributions of model parameter estimates (including the modified bootstrap distributions of the estimated variance components) are then centred at the original estimate values, using a mean correction for regression coefficients, i.e.  $(\hat{\beta}_k^{**}) = \left[ \hat{\beta}_k \mathbf{1}_B + (\hat{\beta}_k^*) - av(\hat{\beta}_k^*) \right]$ , and a ratio correction for variance components, i.e.  $(\hat{\sigma}_u^{2**}) = (\hat{\sigma}_u^{*\text{mod}2}) \times \hat{\sigma}_u^2 \left\{ av(\hat{\sigma}_u^{*\text{mod}2}) \right\}^{-1}$  and  $(\hat{\sigma}_e^{2**}) = (\hat{\sigma}_e^{*\text{mod}2}) \times \hat{\sigma}_e^2 \left\{ av(\hat{\sigma}_e^{*\text{mod}2}) \right\}^{-1}$ . Note that we use a '\*\*' superscript here to distinguish the values defining these adjusted bootstrap distributions from the original bootstrap values generated by SBB, which are denoted by a '\*' superscript.

We refer to the first additional step above as 'tilting' and to the second as 'tethering'. Tilting and tethering together represent a posterior adjustment to the bootstrap distributions generated by SBB that is, as we show in the next subsection, another way of modifying SBB to ensure its consistency under a linear mixed model. Note that bootstrap distributions for quantities that depend on model parameters (e.g. EBLUPs) need to be recomputed using these tilted and tethered bootstrap parameter values. Bootstrap confidence intervals are then defined using the relevant tilted and tethered bootstrap distributions. Because these adjustments are carried out after SBB, we refer to this method as SBB.Post in what follows.

### 2.3.4 Consistency of SBB.Prior and SBB.Post

The semiparametric block bootstrap and its two variations described above are covered by the random effects bootstrap framework described in Field and Welsh (2007). These authors show that the random effects bootstrap gives asymptotically consistent results for the corresponding random effects model under joint asymptotics, i.e. when the number of clusters and the number of observations in each cluster increases. Assuming certain regularity conditions, Shao et al. (2000) show that bootstrap percentile confidence intervals are asymptotically consistent. Carpenter et al. (2003) use the same arguments as in Shao et al. (2000) to prove the asymptotic consistency of CGR-generated bootstrap percentile confidence intervals under the random effects model. This follows from showing that the bootstrap expectations of the ML estimating equations are zero. We now show that this result also holds for SBB.Prior and SBB.Post.

Consider the case of ML estimation under (1), where, up to an additive constant, the log-likelihood function is  $l = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) - \log |\mathbf{V}|$  and  $\mathbf{V}$  is the variance-covariance matrix of  $\mathbf{y}$ . Differentiating this log-likelihood with respect to  $\boldsymbol{\beta}$  leads to the ML estimating function for  $\boldsymbol{\beta}$ , i.e.  $sc(\boldsymbol{\beta}) = \mathbf{x}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})$ . Since  $E^*(y^* - \mathbf{x}\hat{\boldsymbol{\beta}}) = 0$  for SBB.Prior, it follows that the expectation of this estimating function with respect to the bootstrap distribution is zero at  $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ . This shows consistency of  $\hat{\boldsymbol{\beta}}^*$  for  $\hat{\boldsymbol{\beta}}$  under SBB.Prior. In order to demonstrate consistency of the bootstrap estimates of the variance components under SBB.Prior, we note that  $(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) = tr\{\mathbf{V}^{-1}(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T\}$ , see McCulloch and Searle (2001, page 301). Put  $\mathbf{R} = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})(\mathbf{y} - \mathbf{x}\boldsymbol{\beta})^T$ . The log-likelihood function can then be expressed as  $l = -tr(\mathbf{V}^{-1}\mathbf{R}) - \log |\mathbf{V}|$ . The first derivative of this log-likelihood with respect to the variance components parameter  $\delta = (\sigma_u^2, \sigma_e^2)$  defines their estimating function,

$sc(\delta) = -tr\left(\frac{\partial \mathbf{V}^{-1}}{\partial \delta} \mathbf{R}\right) + tr\left(\mathbf{V} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right)$ , so  $E\{sc(\delta)\} = \mathbf{0}$ . Put  $\mathbf{R}^* = (\mathbf{y}^* - \mathbf{x}\hat{\boldsymbol{\beta}})(\mathbf{y}^* - \mathbf{x}\hat{\boldsymbol{\beta}})^T$  and note

that  $\hat{\mathbf{V}} = E^*(\mathbf{R}^*)$ , where  $\hat{\mathbf{V}}$  is the ML estimate of  $V$ . We then need to show that

$E^*\{sc^*(\delta)\} = \mathbf{0}$ , where  $sc^*(\delta) = -tr\left(\frac{\partial \mathbf{V}^{-1}}{\partial \delta} \mathbf{R}^*\right) + tr\left(\mathbf{V} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right)$ . This follows because

$$\begin{aligned} -E^*\left[tr\left(\frac{\partial \mathbf{V}^{-1}}{\partial \delta} \mathbf{R}^*\right)\right] + tr\left(\mathbf{V} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right) &= -E^*\left[tr\left(\mathbf{R}^{*T} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right)\right] + tr\left(\mathbf{V} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right) \\ &= -tr\left(E^*(\mathbf{R}^{*T}) \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right) + tr\left(\mathbf{V} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right) \\ &= -tr\left(\hat{\mathbf{V}} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right) + tr\left(\mathbf{V} \frac{\partial \mathbf{V}^{-1}}{\partial \delta}\right) \\ &\approx 0, \end{aligned}$$

where the last approximate equality is a consequence of the fact that  $\hat{\mathbf{V}}$  and  $\mathbf{V}$  are symmetric and  $\hat{\mathbf{V}} \sim \mathbf{V}$ , where  $\sim$  denotes 'asymptotically equal'. That is,  $\hat{\boldsymbol{\delta}}^* = (\hat{\sigma}_u^{2*}, \hat{\sigma}_e^{2*})$  is consistent for  $\boldsymbol{\delta} = (\sigma_u^2, \sigma_e^2)$  under SBB.Prior. Since ML and REML estimates are asymptotically identical, these consistency properties also hold for REML estimation.

Similar consistency results hold for SBB.Post, since tethering is another way of achieving the same asymptotic behaviour that centering and rescaling guarantees for SBB.Prior and CGR. To show this, we use a superscript of "\*\*\*" to denote post-tethering bootstrap realisations, with  $E^{**}$  denoting the corresponding expectation. Then

$$\begin{aligned} E^{**}(y^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}) &= E^{**}\left\{y^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**} + \mathbf{x}(\hat{\boldsymbol{\beta}}^{**} - \hat{\boldsymbol{\beta}})\right\} \\ &= E^*(y^* - \mathbf{x}\hat{\boldsymbol{\beta}}^*) + \mathbf{x}E^{**}(\hat{\boldsymbol{\beta}}^{**} - \hat{\boldsymbol{\beta}}) \\ &= 0 \end{aligned}$$

since under tethering  $E^{**}(\hat{\boldsymbol{\beta}}^{**}) = \hat{\boldsymbol{\beta}}$ , while the tethered residuals  $y^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**}$  and the untethered residuals  $y^* - \mathbf{x}\hat{\boldsymbol{\beta}}^*$  are identical and  $E^*(y^* - \mathbf{x}\hat{\boldsymbol{\beta}}^*) = 0$ . It immediately follows that the tethered SBB bootstrap is consistent for  $\hat{\boldsymbol{\beta}}$ . To prove the corresponding consistency of this bootstrap

for the estimated variance components, we show that  $\hat{\mathbf{V}} \sim E^{**}(\mathbf{R}^{**})$ , where  $\mathbf{R}^{**} = (\mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}})(\mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}})^T$ . This follows because we can write

$$\begin{aligned} E^{**}(\mathbf{R}^{**}) &= E^{**} \left\{ \mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**} + \mathbf{x}(\hat{\boldsymbol{\beta}}^{**} - \hat{\boldsymbol{\beta}}) \right\} \left\{ \mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**} + \mathbf{x}(\hat{\boldsymbol{\beta}}^{**} - \hat{\boldsymbol{\beta}}) \right\}^T \\ &= \hat{\mathbf{V}} + \mathbf{x} \text{Var}^{**}(\hat{\boldsymbol{\beta}}^{**}) \mathbf{x}^T \end{aligned}$$

where the last equality is a consequence of the fact that under tethering,

$$E^{**} \left\{ \mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**} \right\} \left\{ \mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**} \right\}^T = \hat{\mathbf{V}}$$

and because of the independence of the bootstrap distributions of  $\hat{\boldsymbol{\beta}}^{**}$  and  $\mathbf{y}^{**} - \mathbf{x}\hat{\boldsymbol{\beta}}^{**}$ . It only remains to note that  $\text{Var}^{**}(\hat{\boldsymbol{\beta}}^{**}) = O(n^{-1})$ .

#### 2.3.4 Calibration to the estimated covariance matrix of the variance components

By construction, the rescaling of residuals underpinning the SBB.Prior method leads to level 1 and level 2 bootstrap residuals with variances that are close to the corresponding variance component estimates. However, this does not mean that the covariance matrix of the bootstrap distribution of these variance components is close to the estimated asymptotic covariance matrix of the variance components estimators. This suggests that we may be able to improve on SBB.Prior by calibrating the empirical covariance matrix of the bootstrap estimates of the variance components generated under this procedure to the ML/REML estimate of the asymptotic covariance matrix of the variance components estimators. This can be achieved by a suitable Cholesky decomposition. However, it is important to note that the performance of this second order calibrated block bootstrap then depends on the accuracy of the estimated asymptotic covariance matrix of the variance components estimators. In the simulations reported in the next section we observed that this extra level of calibration lead to undesirable sensitivity to model assumptions. This was not unexpected since this second order

calibration depends on the model (1) being true. Results for this method are therefore not reported, but can be obtained from the authors.

### 3. EMPIRICAL EVALUATIONS

#### 3.1 DESCRIPTION OF THE SIMULATION EXPERIMENTS

We now describe a series of simulation experiments that were used to evaluate the performance of the different bootstrap methods described in the previous Section and which are set out in Table 1. In the first two of these experiments, referred to as simulation sets *A* and *B* below, we used the standard random effects model (1) to generate clustered data. In particular, in both we generated data using a two-level model of the form  $y_{ij} = 1 + 2x_{ij} + u_i + e_{ij}$ ,  $i = 1, \dots, D$ ;  $j = 1, \dots, n_i$ . We fixed the total number of clusters at  $D = 100$  and within each experiment simulated data for two sets of equal cluster sizes,  $n_i = 5$  and  $n_i = 20$ . Values of  $x_{ij}$  were generated independently as  $x_{ij} \sim \text{Uniform}(0,1)$ . The cluster specific (level 2) random errors  $u_i$  and the individual specific (level 1) random errors  $e_{ij}$  were generated as mutually independent and identically distributed random variables with zero means and with variances  $\sigma_u^2$  and  $\sigma_e^2$  respectively.

In simulation set *A*,  $u_i \sim N(0, \sigma_u^2 = 0.04)$  and  $e_{ij} \sim N(0, \sigma_e^2 = 0.16)$ . In simulation set *B*, we generated  $u_i$  from a  $\chi^2$  distribution with mean zero and variance  $\sigma_u^2 = 0.04$  as  $u_i \sim 0.2 \left[ (\chi_1^2 - 1) / \sqrt{2} \right]$ . Similarly, we generated the individual level errors  $e_{ij}$  independently of the cluster level errors  $u_i$  from a  $\chi^2$  distribution with mean zero and variance  $\sigma_e^2 = 0.16$  as  $e_{ij} \sim 0.4 \left[ (\chi_1^2 - 1) / \sqrt{2} \right]$ .

Note that in both set *A* and set *B*, units within a cluster are equi-correlated. Since our interest is in clustered data situations where this does not hold, we investigated an alternative

to set  $A$  where the individual level errors  $e_{ij}$  were generated so that within cluster units are not equi-correlated. In this case individual level errors within a cluster were simulated so that they corresponded to a first order auto-correlated series of form  $e_{ij} = \lambda e_{i(j-1)} + \varepsilon_{ij}, j = 1, \dots, n_i$  with  $\lambda = 0.5$  and  $\varepsilon_{ij} \sim N(0,1)$ . This is referred to as simulation set  $C$  below. Finally, we investigated the impact of correlation between units in different clusters in a fourth set of simulations, denoted simulation set  $D$  below, where we replicated simulation set  $C$  except that all individual level errors were now generated from the same first order auto-correlated series of size  $n = \sum_{i=1}^D n_i$  as  $e_{ij} = \lambda e_{i(j-1)} + \varepsilon_{ij}, j = 1, \dots, n$ . This simulation therefore approximates the type of time series problem that motivated the development of the block bootstrap.

A total of  $R = 1000$  Monte Carlo simulations were carried out for each simulation set, and within each simulation we implemented each of the bootstrap methods set out in Table 1 using  $B = 1000$  bootstrap replicates. This number of simulations and bootstrap samples is suitable for evaluating 95 per cent percentile confidence intervals, see Caers et al. (1998).

### 3.2 DISCUSSION OF SIMULATION RESULTS

Average coverage rates of nominal 95 percent bootstrap confidence intervals for the various model parameters were obtained for the different simulations sets. These coverage rates are reported in Table 2. It is clear that there is not much difference in the coverage rates for the regression coefficients (i.e.  $\alpha$  and  $\beta$ ) between the different bootstrap methods and between the different simulation sets, with the notable exception that the CGR method recorded low coverage for  $\alpha$  in our large cluster size ( $n_i = 20$ ) simulations, indicating a potential bias problem with our implementation of this method.

It is well known that classical estimation inference for the variance component parameters of (1) are sensitive to deviations from this model. As a consequence we now focus



on bootstrap coverage performance for the variance component parameters  $\sigma_u^2$  and  $\sigma_e^2$ . In simulation set *A* the assumed model is true, and so the coverage rates of the parametric bootstrap and CGR are around 95 per cent. In contrast, SBB records low coverage, especially for small ( $n_i = 5$ ) cluster sizes. This is effectively corrected by SBB.Post and SBB.Prior, although there is evidence that for small cluster sizes SBB.Prior is more effective than SBB.Post.

Turning to results from simulation set *B*, we see that the parametric bootstrap fails. The performance of CGR is better, but is still unsatisfactory. In contrast, although the simple block bootstrap SBB is remains unsatisfactory for small cluster sizes, its performance for large cluster sizes is good. This performance is reversed for SBB.Prior which performs better for small cluster sizes than for large cluster sizes. Cluster size does not seem to impact as much on SBB.Post, which performs adequately and seems better than CGR in this simulation.

The performances of the different bootstrap methods in sets *C* and *D* were qualitatively similar to those recorded for sets *A* and *B*. The simple block bootstrap SBB fails when cluster sizes are small and recovers somewhat as the cluster size increases. The performances of both the parametric (Para) and CGR bootstraps are on a par, as are those of SBB.Post and SBB.Prior, with SBB.Prior the better performer for small cluster sizes. Overall, SBB.Prior appears to be the best performing of the five bootstrap methods that we investigated, with SBB.Post a little behind. Both these bootstrap methods seem robust to the departures from model assumptions that we considered in our simulations.

Although we do not present these results here, we also carried out number of simulation studies that examined the performance of the bootstrap methods set out in Table 1 in other situations, all of which have some relevance to real life data:

- We replicated simulation sets *A* to *C* with a smaller number of clusters, i.e.  $D = 50$ , and noted that the relative performances of the different bootstrap methods were almost identical to those observed when  $D = 100$ .
- We examined the impact of misspecification of the cluster structure in the block bootstrap by replicating simulation set *A* with data generation and model fitting based on  $D = 100$  clusters, but with bootstrap data generated using a smaller number  $D = 50$  clusters. This did not change the behaviour of the block bootstrap methods.
- We also examined the impact of varying cluster sizes by replicating simulation set *A* with cluster sizes ranging from 1 to 100, with about half the clusters having 10 or fewer observations. Again, the block bootstrap methods SBB.Post and SBB.Prior performed satisfactorily.

#### **4. APPLICATION TO ENVIRONMENTAL DATA MODELLING**

In this Section we apply the different bootstrap methods set out in Table 1 to the environmental data analysed in Beare et al. (2011). These data consist of  $n = 3177$  values of positive daily rainfall measured at a group of rain gauges over a period of approximately four months, together with the values of 37 covariates measuring daily meteorological conditions as well as the spatial characteristics of the different gauges. The data were collected as part of a trial of the effect of two ground-based cloud ionizing devices on downwind rainfall, and so the covariates include measurements relating to the daily operational status of the two devices as well as the distance and downwind orientation of a gauge relative to each device on a day. Since the hypothesised impact of these devices is to enhance downwind rainfall, it is necessary to include terms in the model for observed rainfall that allow for the natural variation in rainfall due to the spatial and temporal inhomogeneity of rain cloud movement

over the target downwind area. In the analysis described in Beare et al. (2011) this was done by including random effects for 397 spatio-temporal clusters in the rainfall model, where these clusters were defined by groups of gauges that had similar relative orientations to the two devices on a day. The distribution of these spatio-temporal cluster sizes can be seen in Figure 1, and we note that they vary from minimum of 1 gauge to maximum of 57 gauges, with average size of 8 gauges.

A more conservative approach to defining cluster random effects for these rain data is based on the fact that the random sequence used to control the operation of the two devices was essentially made up of 4 day 'blocks'. Assuming that there could be significant unexplained between block and between device heterogeneity in rainfall then leads one to consider random spatio-temporal effects based on clustering gauge-day rainfall measurements by both 4 day block and whether the rainfall measurement is for a gauge that is downwind of only one of the devices or downwind of both. We refer to these clusters as 4 day downwind clusters in what follows. There are 83 such clusters in the data, and the distribution of their corresponding sizes is shown in Figure 1. Note that these sizes range from 1 to 197 with average of 38.

The next issue that needs to be addressed is the scale at which the daily rainfall data is modelled using (1). Clearly, we can fit this model to the actual rainfall values. However, given that rainfall measurements are strictly positive and heavily skewed, an obvious alternative is use (1) as a model for the logarithm of rainfall. The marginal distributions of daily rainfall on the raw scale and on the log scale are shown in Figure 2. The apparently discrete nature of the distribution of log rainfall for small values of this variable evident in Figure 2 is due to the fact that rainfall in gauges is measured in increments of 0.2 mm. Figure 3 allows one to compare the predicted values (i.e. fitted values for fixed effects plus predicted random effects) generated by fitting (1) to both raw rainfall as well as to log rainfall using

both spatio-temporal clusters and 4 day downwind clusters, with a REML fit used in all cases. This clearly shows that fitting (1) to log rainfall is the better choice. It also demonstrates that a random effects specification using spatio-temporal clusters leads to a better fit than a random effects specification using 4 day downwind clusters.

In Section 3 we noted that the block bootstrap methods SBB.Post and SBB.Prior should be robust to the assumption that level 1 and level 2 errors in (1) are independent and identically distributed Gaussian variables. Although this assumption may be reasonable when (1) is fitted to log rainfall, it is clear from Figure 3 that it is hard to justify when (1) is fitted using raw rainfall values. We therefore examine the application of bootstrap methods to the rainfall data under both types of clusters as well as when (1) is fitted to raw rainfall and to log rainfall. This leads to 4 sets of analyses. These are reported in Table 3 and in Figures 4 and 5.

Our initial analysis focussed on comparing the bootstrap tests of significance for the fixed effects in the model, where we decided that an effect is significant if its 95 per cent confidence interval does not include zero. In no case did we observe a situation where the standard parametric test (i.e. one based on the asymptotic REML-based confidence interval) led to a different conclusion about significance compared with any of the bootstrap tests. This is consistent with the results that we obtained in our simulations, and so we do not show them here. They can be obtained from the authors on request.

However, we did observe substantial differences between the different bootstrap methods as far as inferences about the variance components in the model are concerned. Table 3 shows the estimated standard errors and Figure 4 shows the associated 95 per cent confidence intervals for these components generated by the different bootstrap methods under the four different model specifications. The corresponding bootstrap sampling distributions for these variance components under these model specifications are shown in Figure 5. We

see that the estimated standard errors generated by the REML fit of (1) (denoted Regression) and those generated by the parametric bootstrap method (denoted Para) are very close (see Table 3). The estimated standard errors generated by the CGR bootstrap are also very close to those generated by Para and by Regression when the model is fitted on log scale. When the model is fitted on the raw scale, these estimated standard errors are larger. However, in all cases the estimated standard errors generated by the block bootstrap methods are much larger (often more than twice as large) as the estimated standard errors generated by CGR, Para and Regression. Since there is considerable doubt about (1) as a model for actual rainfall values, plus concern about the validity of the homogeneous random effects assumptions when (1) is fitted on the log scale, these results imply that the more conservative estimated standard errors generated by the block bootstrap methods may be preferable. This conclusion is reinforced by the confidence intervals displayed in Figure 4. These show that the intervals defined by Regression, Para and CGR are qualitatively very similar, and typically narrower than those generated by SBB.Post and SBB.Prior. They also show that the intervals generated by the unmodified block bootstrap SBB tend to be biased upwards in the case  $\sigma_u^2$  and biased downwards in the case of  $\sigma_e^2$ . Of more concern, however, is the extreme narrowness of the intervals for  $\sigma_e^2$  generated by Regression, Para and CGR. This concern is reinforced when we examine the bootstrap distributions for these methods shown in Figure 5, which appear to show unwarranted precision as far as estimation of the variance components in the model is concerned. In contrast, the bootstrap distributions generated by SBB.Post and SBB.Prior appear more realistic. These conclusions are consistent with the conclusions drawn from the simulations described in Section 3 where we noted that in case of non-normal data, both Para and CGR lead to under coverage, while both SBB.Post and SBB.Prior lead to intervals with coverage that is much closer to nominal levels.

## 5. CONCLUDING REMARKS

Our aim in this paper has been to describe and to evaluate an alternative semiparametric block bootstrap method for clustered data. The method itself is described in Section 2, and, when used with either additional post-bootstrap processing (SBB.Post) or with modified nonparametric level 1 and level 2 residuals (SBB.Prior), appears to provide a simple and robust alternative to the model dependent bootstrap methods for clustered data that are presently available in the literature. Given that the first order structure of the underlying linear mixed model is adequately specified, both SBB.Post and SBB.Prior account for within cluster heterogeneity as well as between cluster dependence. This good performance is demonstrated in the application to an environmental data set in Section 3, where we observe that it was only these block bootstrap methods that provided realistic results across all four modelling scenarios that we investigated.

Extension of SBB.Prior and SBB.Post to versions of (1) that include random slope parameters is straightforward. We let  $\mathbf{z}_{ij}$  be a  $q \times 1$  vector of group level covariates for unit  $j$  in cluster  $i$  and replace model (1) by

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T u_i + e_{ij}. \quad (2)$$

The only change to SBB that is required in this case is the definition of the level 2 average residual  $\bar{r}_h$  for cluster  $h$ . This can be replaced by the  $q \times 1$  vector of level 2 average residuals for group  $h$ :  $\mathbf{r}_h^{(2)} = (\mathbf{z}_h^T \mathbf{z}_h)^{-1} \mathbf{z}_h^T \mathbf{r}_h$ ,  $h = 1, \dots, D$ , where  $\mathbf{z}_h$  is the  $n_h \times q$  matrix of  $\mathbf{z}_{ij}$  for group  $h$  and  $\mathbf{r}_h$  is  $n_h \times 1$  vector of marginal residuals. Investigation of the empirical performance of this extension is currently under way, as is research into extending SBB.Prior and SBB.Post to generalised linear mixed models, and to M-quantile-based alternatives (Chambers and Tzavidis, 2006) to (2) above.

## REFERENCES

- Beare, S., Chambers, R., Peak, S. and Ring, J.M. (2011), "Accounting for Spatiotemporal Variation of Rainfall Measurements When Evaluating Ground-based Methods of Weather Modification," *Journal of Weather Modification*, 43, 44-63.
- Caers, J., Beirlant, J. and Vynckier, P. (1998), "Bootstrap Confidence Intervals for Tail Indices," *Computational Statistics and Data Analysis*, 26, 259-277.
- Canty, A.J. and Davison, A.C. (1999), "Resampling-based Variance Estimation for Labour Force Surveys," *Statistician*, 48, 379–391.
- Carpenter, J.R., Goldstein, H. and Rasbash, J. (2003), "A Novel Bootstrap Procedure for Assessing the Relationship Between Class Size and Achievement," *Journal of the Royal Statistical Society, Series C*, 52, 431-443.
- Chambers, R. and Tzavidis, N. (2006), "M-quantile Models for Small Area Estimation," *Biometrika*, 93, 255-268.
- Clark, R.G. and Allingham, S. (2011), "Robust Resampling Confidence Intervals for Empirical Variograms," *Mathematical Geosciences*, 43(2), 243-259.
- Davison, A.C. and Hinkley, D.V. (1997), *Bootstrap Methods and Their Application*, Cambridge: Cambridge University Press.
- DiCiccio, T.J. and B. Efron. (1996), "Bootstrap Confidence Intervals," *Statistical Science*, 11(3), 189–212.
- Diciccio, T.J. and Romano, J.P. (1988), "A Review of Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society, Series B*, 50, 338-354.
- Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.
- Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall.

- Field, C.A., and Welsh, A.H. (2007), "Bootstrapping Clustered Data," *Journal of the Royal Statistical Society, Series B*, 69, 369–390.
- Hall, P., Horowitz, J.L. and Jing, B. (1995), "On Blocking Rule for the Bootstrap with Dependent Data," *Biometrika*, 82 (3), 561-574.
- Harville, D. (1977), "The Use of Linear-Model Methodology to Rate High School or College Football Teams," *Journal of the American Statistical Association*, 72, 278-289.
- Hutson, A.D. (2004), "A Semiparametric Bootstrap Approach to Correlated Data Analysis Problems," *Computer Methods and Programs in Biomedicine*, 73, 129-134.
- McCulloch, C.E. and Searle, R.S. (2001), *Generalized, Linear, and Mixed Models*, New York: John Wiley & Sons, Inc.
- Rasbash, J., Browne, W, Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I. and Lewis, T. (2000), "A *User's Guide to MLwiN*," London: Institute of Education, London.
- Rao, J.N.K. and Wu, C. F. J. (1988), "Resampling Inference With Complex Survey Data," *Journal of American Statistical Association*, 83, 231–241.
- Shao, J., Kübler, J. and Pigeot, I. (2000), "Consistency of the Bootstrap Procedure in Individual Bioequivalence," *Biometrika*, 87, 573-585.
- Shao, J. and Tu, D. (1995), *The Jackknife and Bootstrap*, New York: Springer, New York.



**Table 1.** Description of bootstrap methods used in the simulation studies.

Name	Description of bootstrap method
Para	Two-level parametric bootstrap
CGR	Carpenter et al. (2003) bootstrap
SBB	Semiparametric block bootstrap
SBB.Prior	Semiparametric block bootstrap with centred and rescaled residuals
SBB.Post	Semiparametric block bootstrap with tilting and tethering adjustments

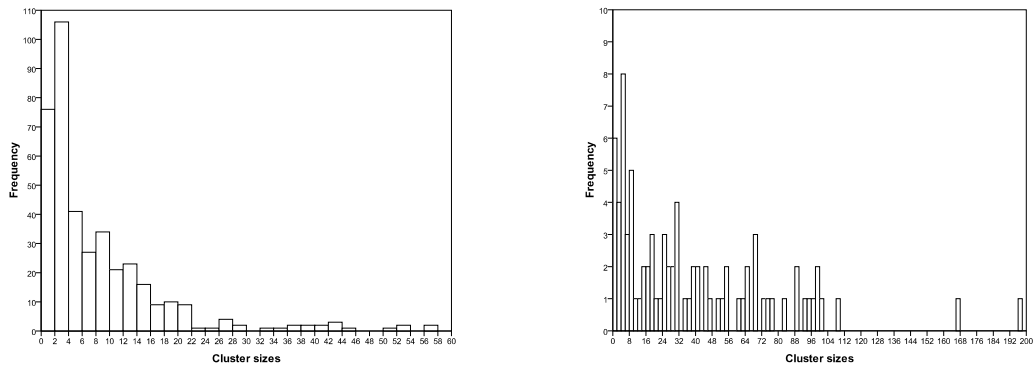
**Table 2.** Average coverage rates of nominal 95 per cent bootstrap confidence intervals for model parameters, simulation sets *A - D*.

Method	$n_i = 5$				$n_i = 20$			
	$\alpha$	$\beta$	$\sigma_u^2$	$\sigma_e^2$	$\alpha$	$\beta$	$\sigma_u^2$	$\sigma_e^2$
Set A								
Para	0.95	0.94	0.95	0.95	0.96	0.95	0.93	0.95
CGR	0.91	0.95	0.95	0.94	0.81	0.95	0.94	0.94
SBB	0.95	0.93	0.23	0.26	0.97	0.95	0.89	0.86
SBB.Prior	0.95	0.95	0.96	0.98	0.96	0.95	0.94	0.99
SBB.Post	0.95	0.93	0.86	0.99	0.97	0.95	0.94	0.99
Set B								
Para	0.95	0.95	0.77	0.59	0.94	0.95	0.61	0.54
CGR	0.93	0.95	0.84	0.87	0.81	0.95	0.82	0.86
SBB	0.96	0.93	0.64	0.72	0.95	0.95	0.94	0.96
SBB.Prior	0.96	0.95	0.94	0.96	0.94	0.95	0.82	0.99
SBB.Post	0.96	0.93	0.82	0.97	0.95	0.95	0.83	0.99
Set C								
Para	0.91	0.96	0.95	0.77	0.94	0.93	0.95	0.82
CGR	0.88	0.94	0.95	0.76	0.77	0.95	0.94	0.83
SBB	0.93	0.92	0.28	0.08	0.95	0.94	0.90	0.64
SBB.Prior	0.92	0.94	0.96	0.89	0.93	0.95	0.94	0.95
SBB.Post	0.93	0.92	0.88	0.93	0.95	0.94	0.93	0.97
Set D								
Para	0.90	0.96	0.94	0.88	0.94	0.95	0.94	0.88
CGR	0.87	0.94	0.94	0.86	0.75	0.95	0.95	0.87
SBB	0.91	0.92	0.24	0.27	0.94	0.93	0.90	0.78
SBB.Prior	0.91	0.94	0.96	0.96	0.93	0.94	0.94	0.95
SBB.Post	0.91	0.92	0.85	0.96	0.94	0.93	0.94	0.95

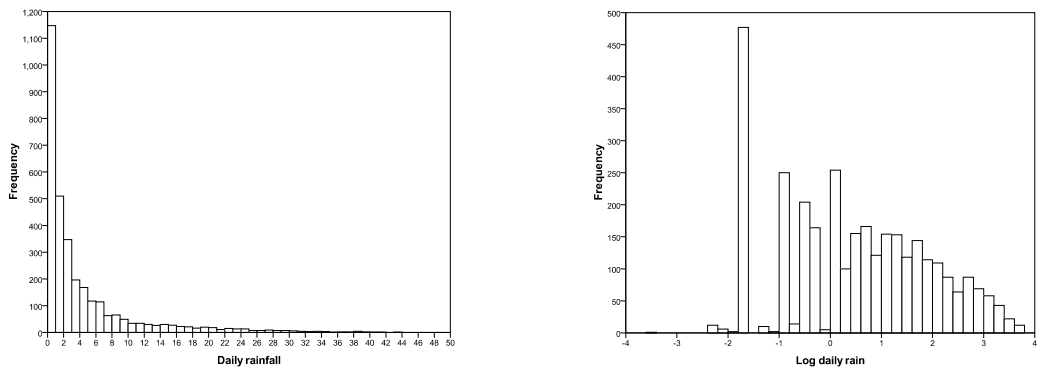
**Table 3.** Bootstrap estimates of standard errors for estimates of variance components for environmental data set. Note that Regression is the estimated asymptotic standard error produced under REML.

Cluster	Model	Parameter	Estimate	Estimated Standard Error					
				Regression	Para	CGR	SBB	SBB.Prior	SBB.Post
Spatio-temporal	Rain	$\sigma_u^2$	5.622	0.610	0.625	0.957	1.366	1.176	0.915
		$\sigma_e^2$	13.207	0.352	0.352	0.701	2.081	2.357	3.172
	Log Rain	$\sigma_u^2$	0.306	0.032	0.033	0.035	0.063	0.043	0.033
		$\sigma_e^2$	0.654	0.017	0.017	0.018	0.054	0.062	0.078
4 day downwind	Rain	$\sigma_u^2$	4.246	0.815	0.862	1.104	1.460	1.202	1.064
		$\sigma_e^2$	15.269	0.388	0.389	0.823	2.800	2.902	3.801
	Log Rain	$\sigma_u^2$	0.206	0.040	0.042	0.036	0.166	0.081	0.078
		$\sigma_e^2$	0.775	0.020	0.020	0.021	0.099	0.103	0.099

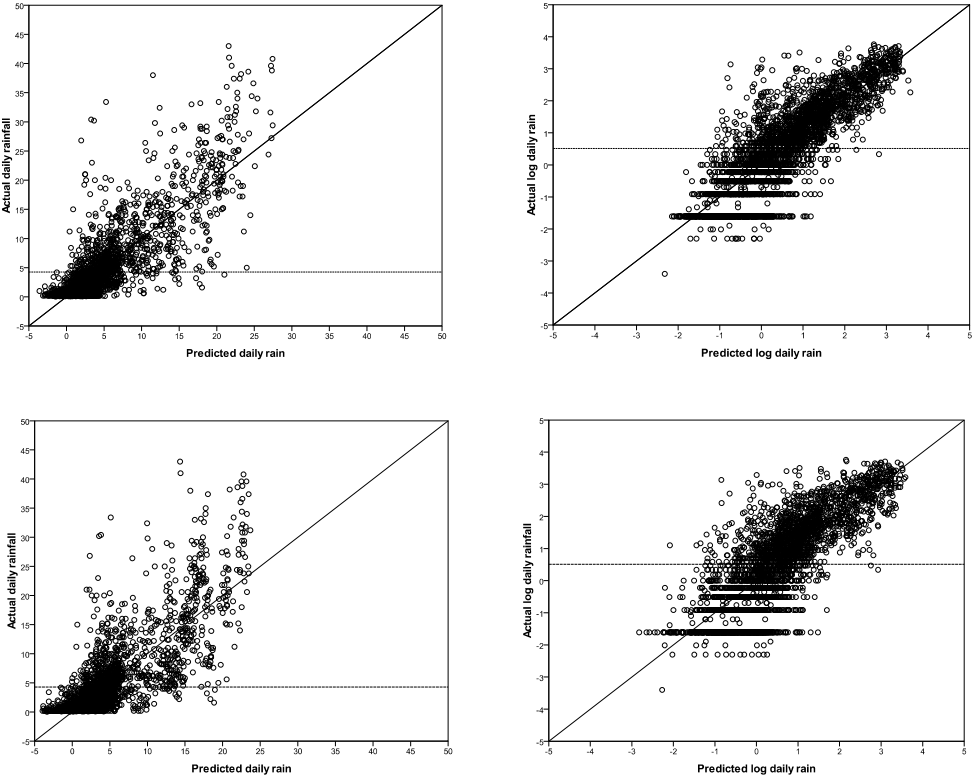
**Figure 1.** Distributions of cluster sizes for environmental data set, with spatio-temporal clusters on the left and 4 day downwind clusters on the right.



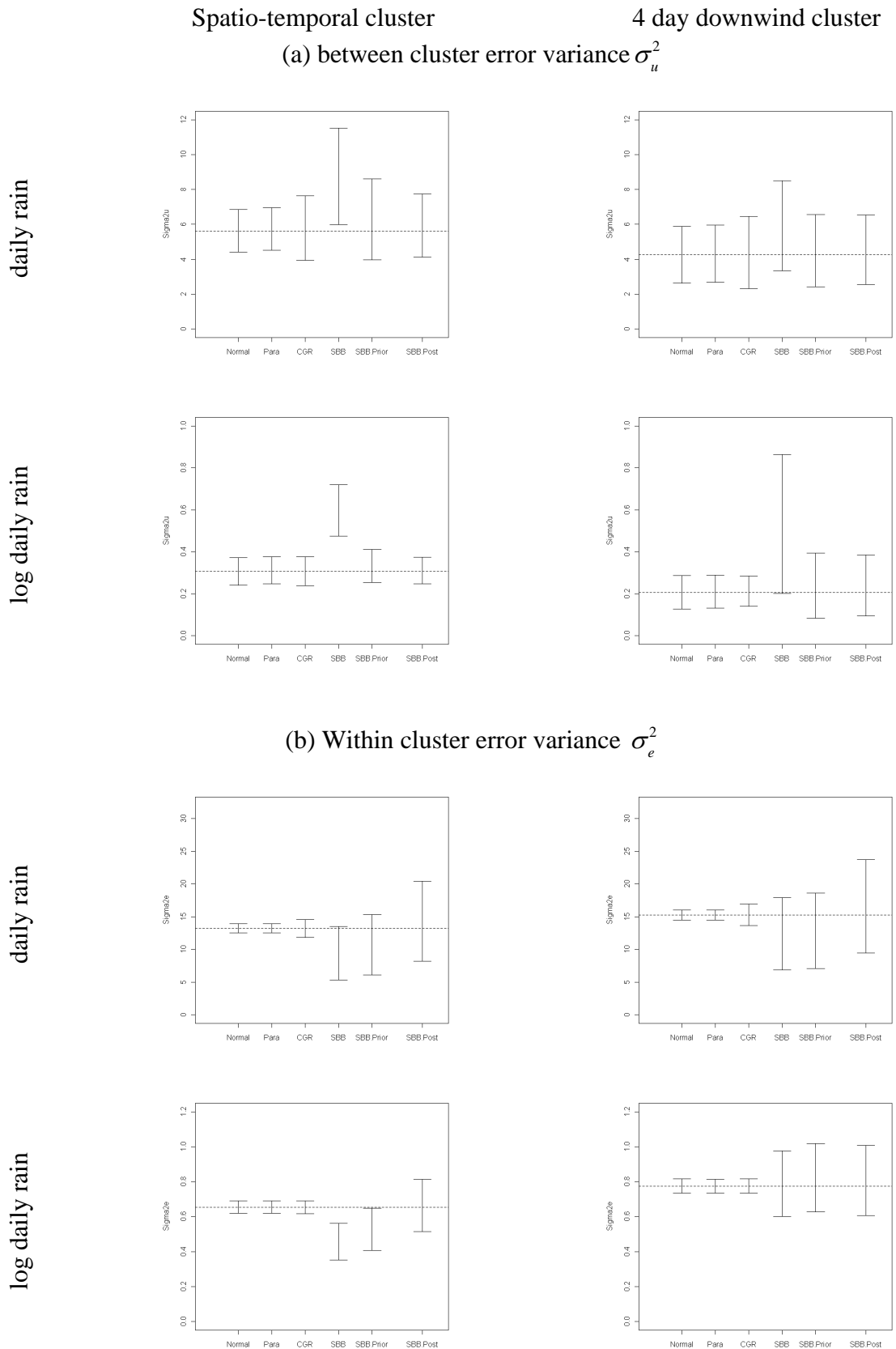
**Figure 2.** Distribution of daily rainfall for environmental data set - raw scale (left) and log scale (right).



**Figure 3.** Plots of actual vs. predicted values for linear mixed model fitted to environmental data set using daily rainfall values (left column) and log daily rainfall values (right column). Solid line is  $y = x$  line and dotted line is average value of  $y$ . Top row corresponds to model with spatio-temporal clusters, while bottom row corresponds to model with 4 day downwind clusters.

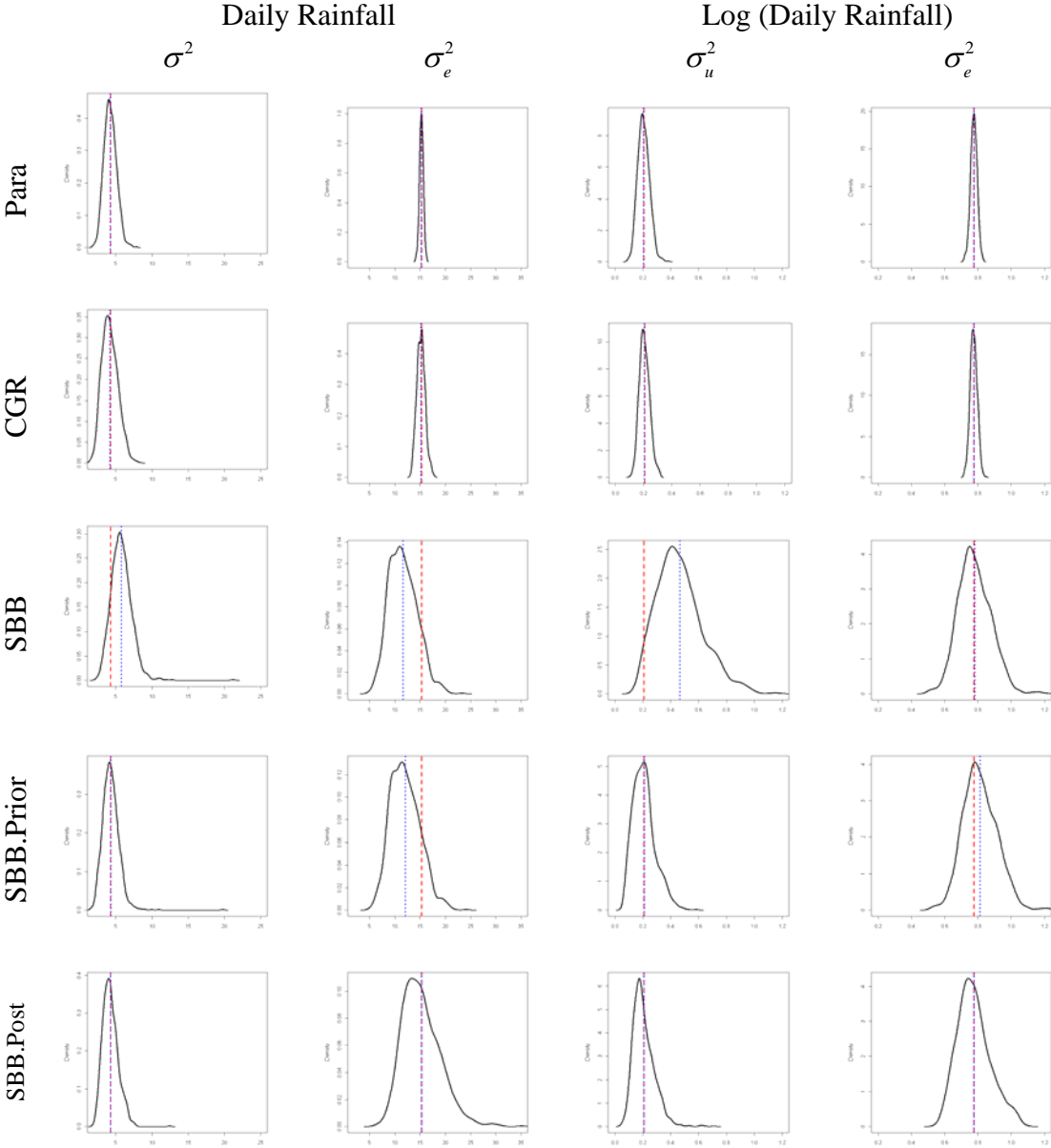


**Figure 4.** Nominal 95 per cent confidence intervals for variance components for environmental data set. Horizontal line in each plot is estimated value of parameter.



**Figure 5.** Bootstrap distributions of estimates of variance components for environmental data set. Dashed line shows the value of the estimate and the dotted line shows the mean of the bootstrap distribution.

(a) 4 day downwind clusters



(b) Spatio-temporal clusters

