J. R. Statist. Soc. B (2007)
69, *Part* 3, *pp.* 369–390

Bootstrapping clustered data

C. A. Field

Dalhousie University, Halifax, Canada

and A. H. Welsh

Australian National University, Canberra, Australia

[Received November 2005. Revised December 2006]

Summary. Various bootstraps have been proposed for bootstrapping clustered data from one-way arrays. The simulation results in the literature suggest that some of these methods work quite well in practice; the theoretical results are limited and more mixed in their conclusions. For example, McCullagh reached negative conclusions about the use of non-parametric bootstraps for one-way arrays. The purpose of this paper is to extend our understanding of the issues by discussing the effect of different ways of modelling clustered data, the criteria for successful bootstraps used in the literature and extending the theory from functions of the sample mean to include functions of the between and within sums of squares and non-parametric bootstraps to include model-based bootstraps. We determine that the consistency of variance estimates for a bootstrap method depends on the choice of model with the residual bootstrap giving consistency under the transformation model whereas the cluster bootstrap gives consistent estimates under both the transformation and the random-effect model. In addition we note that the criteria based on the distribution of the bootstrap observations are not really useful in assessing consistency.

Keywords: Between and within sums of squares; Bootstrap; Clusters; Hierarchical data; One-way arrays

1. Introduction

Our aim in this paper is to address some issues which arise when bootstrapping clustered, hierarchical or multilevel data. These issues arise from the dependence structure in the data, different ways of describing or modelling the data, different ways of bootstrapping and different ways of evaluating the bootstrap.

The problem of bootstrapping clustered data was first discussed in the design-based finite population survey literature by Gross (1980), McCarthy and Snowden (1985), Rao and Wu (1988), Sitter (1992) and others; see Canty and Davison (1999) for a recent application and Shao (2003) for a recent review. In this work, the observations are fixed and inferences are made with respect to the random-sampling mechanism, so the main issue is accommodating complex survey designs arising from various forms of cluster sampling into inference about finite population parameters. The infinite population version of the problem is of quite a different nature and has led to different bootstrap methods; see for example Davison and Hinkley (1997), pages 100–102, McCullagh (2000), Andersson and Karlsson (2001), Ukoumunne *et al.* (2003) and Carpenter *et al.* (2003). Here, the clusters are treated as arising from the data-generating mechanism rather than the sampling scheme and the focus is on making inferences about parameters

Address for correspondence: A. H. Welsh, Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia.
E-mail: Alan.Welsh@anu.edu.au

 1369–7412/07/69369

while accommodating the dependence structure in the data. The infinite population approach is useful in and has been applied to model-based finite population analysis; see Butar and Lahiri (2003), Lahiri (2003) and Welsh (2005).

We shall work in the infinite population context so that inference is with respect to the population model rather than the sampling mechanism. This means that it is helpful to specify a model for how we think the data were generated. We specify two possible models for the data (called the random-effect model and the transformation model). These models are compatible in the sense that their first and second moments agree but, in general, their higher moments do not. This means that the variances of statistics like the sums of squares which depend on fourth or higher moments of the underlying distribution can be different under the two models. In addition, bootstrap methods can have different properties under the two models. This paper explores the consequences of this important fact.

There is already a rich class of bootstrap methods for clustered data in the literature but there is an absence of detailed theoretical results on the properties of these methods. We therefore concentrate on gaining insight into bootstrapping clustered data by deriving theoretical results for methods which are already in the literature.

Davison and Hinkley (1997), pages 100–102, discussed the *randomized cluster bootstrap* ('Strategy 1') in which clusters are selected by simple random sampling with replacement and then observations within clusters are randomly permuted, the *two-stage bootstrap* ('Strategy 2') in which, after selecting clusters as in the randomized cluster bootstrap, observations within clusters are selected by simple random sampling with replacement, and the *random-effect bootstrap* in which we compute predictors of the random effects in the model, draw bootstrap samples of these predictors by simple random sampling with replacement from each of the sets of predictors and then construct bootstrap observations by combining the bootstrap predictors appropriately. The *cluster bootstrap* is a simplified version of the randomized cluster bootstrap in which clusters are selected by simple random sampling with replacement and there is no subsequent permutation. McCullagh (2000) considered the randomized cluster and the two-stage bootstraps as well as the *reverse two-stage bootstrap* in which observations within clusters are first selected by simple random sampling with replacement and then clusters are selected by simple random sampling with replacement. (The reverse two-stage bootstrap has two levels of variability like the two-stage bootstrap but differs from it in that clusters which are selected more than once must contain the same observations each time.) He also considered other non-parametric bootstraps which involved sampling without replacement or permutation. Andersson and Karlsson (2001) considered, among others, the random-effects bootstrap and the *residual bootstrap* in which we compute appropriate residuals, draw a bootstrap sample of these residuals by simple random sampling with replacement from the residuals and then reconstruct bootstrap observations from the bootstrap residuals. Subsequent work has used these bootstraps (Ukoumunne *et al.* (2003) used the cluster bootstrap and Carpenter *et al.* (2003) the random-effect bootstrap) or parametric versions of them (Butar and Lahiri, 2003; Lahiri, 2003). All these bootstraps can be viewed as constructing empirical estimates of the distribution function of the data which can then be used in place of the distribution that actually generated the data to make inferences.

We consider the simplest infinite population problem in which the clustered data can be arranged in a balanced single classification and establish base-line theoretical results for this case. In spite of the simplicity of this set-up, most of the important issues are already present. We also focus on the variance parameters rather than the mean parameters. Functions of the variance components are often of more interest than functions of the mean parameters. Consequently we expand the class of statistics to include those which are functions of the between and within sums of squares, including various variance component estimates, estimates of ratios of variance components and estimates of the intracluster correlation (which were considered by Ukoumunne *et al.* (2003)) or heritability. The properties of all such estimators can be derived from those of the between and within sums of squares, so we derive and present these properties, including those of the sample total for completeness.

The existing literature uses a range of criteria to evaluate bootstrap methods for clustered data; for general discussion see Shao and Tu (1995). Davison and Hinkley (1997), focusing on inference for the mean, argued that the bootstrap observations should mimic the first two moment structure of the original observations as closely as possible. Specifically, they compared the expected first two bootstrap moments of the bootstrap observations with the first two moments of the observations. They concluded that the randomized cluster bootstrap and the random-effect bootstrap (with appropriately chosen predictors of the random effects) most closely mimic the first two moment structure of the observations. McCullagh (2000) argued that the bootstrap distribution should be exchangeable under a suitable group structure. He went on to show that not all bootstraps which meet this criterion produce consistent estimates of the variance of the sample mean or total. This means that, even under his restrictive definition of a bootstrap (he excluded the random-effect and residual bootstraps and his definition does not apply to unbalanced problems), his criterion does not exclude bootstraps which fail. Andersson and Karlsson (2001) used a simulation study to investigate the size and power of bootstrap tests of a hypothesis on a slope parameter and obtained good results for their bootstraps. Similarly, Ukoumunne *et al.* (2003) and Carpenter *et al.* (2003) obtained good results in simulation studies to evaluate the properties of bootstrap confidence intervals for the intracluster correlation coefficient and for regression parameters and components of variance respectively. The work of Ukoumunne *et al.* (2003) shows the importance of using an appropriate transformation when constructing confidence intervals for functions of the variance components. There is therefore already a good body of simulation results which we supplement in this paper with detailed theoretical insight.

We adopt the viewpoint that consistent estimation of the variance of a range of statistics is a minimal requirement for a bootstrap to be useful. We can consider consistency under various asymptotic situations: *cluster asymptotics* in which the number of clusters increases, *observation asymptotics* in which the number of observations in each of the fixed set of clusters increases and *joint asymptotics* in which the number of clusters and the number of observations in each cluster increase. Not all the variance parameters in the model can be estimated consistently under observation asymptotics, so we do not consider this approach further; cluster asymptotics require less of the sequence than joint asymptotics and provide a more stringent criterion in the sense that consistency under cluster asymptotics will give consistency under joint asymptotics.

The statistics that we consider are invariant to permutation of the clusters and/or of observations within clusters. This means that the results for the randomized cluster bootstrap and the cluster bootstrap (which is the same as the randomized cluster bootstrap without the permutation step) are identical. However, the observations themselves are not invariant to permutation of the observations within clusters and so have different first two moment structures (which have different expectations) under the randomized cluster bootstrap and the cluster bootstrap. This observation undermines the value of the moment matching criterion of Davison and Hinkley (1997). Indeed, we argue that it is unclear how to implement this criterion and that matching the moments of the observations is not sufficiently sophisticated to distinguish between bootstrap methods. This is not altogether surprising as we consider statistics (like the sums of squares) which are non-linear functions of the observations and have different properties under different models.

We show that the random-effect bootstrap gives asymptotically consistent results for the random-effect model under joint asymptotics and the residual bootstrap gives asymptotically consistent results for the transformation model under cluster asymptotics. These methods have the advantage that they generalize naturally to more complex random or mixed models but the disadvantage that they are tied to particular models. We also show that the randomized cluster bootstrap and the cluster bootstrap are asymptotically consistent for both models under cluster asymptotics and are the best of the exchangeable bootstraps.

In the next section, we introduce our notation and obtain the results that we require to assess the various bootstrap methods. These results give the second-order moment structure of the sums of squares for non-normal data and compare them under the two different models that we consider. In Section 3, we obtain the first two bootstrap moments of bootstrap versions of the sums of squares under various bootstraps and compare these with the results that are obtained in Section 2. We present simulation results in Section 4 and our conclusions in Section 5.

2. Notation and basic moment results

Consider a balanced single classification with g clusters (groups or blocks) of m observations each. Let Y_{ij} denote the jth observation in the ith cluster, let $y_i^{\mathrm{T}} = (Y_{i1}, \ldots, Y_{im})$ denote the m-vector of observations from the ith cluster and let $y^{\mathrm{T}} = (y_1^{\mathrm{T}}, \ldots, y_g^{\mathrm{T}})$ denote the vector of all mg observations. Then let

$$\bar{Y}_{i\cdot} = m^{-1} \sum_{j=1}^{m} Y_{ij}$$

denote the mean of the ith cluster and

$$\bar{Y}_{\cdot\cdot} = (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} Y_{ij} = g^{-1} \sum_{i=1}^{g} \bar{Y}_{i\cdot}$$

denote the grand mean. Next, define the sample total

$$T = mg\bar{Y}_{\cdot\cdot} = m \sum_{i=1}^{g} \bar{Y}_{i\cdot}$$

and the between- and within-cluster statistics

$$S_{\mathrm{B}k} = m \sum_{i=1}^{g} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^k$$

and

$$S_{\mathrm{W}k} = \sum_{i=1}^{g} S_{\mathrm{W}ki},$$

where

$$S_{\mathrm{W}ki} = \sum_{j=1}^{m} (Y_{ij} - \bar{Y}_{i\cdot})^k, \qquad k = 2, 3, 4.$$

These statistics with $k = 2$ are the between-cluster sum of squares $S_{\mathrm{B}2}$ and the within-cluster sum of squares $S_{\mathrm{W}2}$.

We shall focus mainly on the between and within sums of squares because many other statistics of interest (including the maximum likelihood estimators under the normal model and the method-of-moments estimators) can be expressed as smooth functions of the sample total and these sums of squares. At least asymptotically, the properties of such statistics are easily derived from those of the total and the two sums of squares. These properties depend on the model and hence on the way that the data were generated. We consider two simple models and explore the properties of the estimators under both models.

The simplest model for a balanced single classification is the *random-effect model*

$$Y_{ij}=\mu+\beta_i+\varepsilon_{ij}, \qquad j=1,\ldots,m, \quad i=1,\ldots,g, \tag{1}$$

where β_i are independent and identically distributed random variables with mean 0 and variance σ_β^2, independent of the independent and identically distributed random variables ε_{ij} which have mean 0 and variance σ_ε^2. We assume that $E(\beta^4)<\infty$ and $E(\varepsilon^4)<\infty$. The observations Y_{ij} and Y_{kl} are independent whenever $i\neq k$ and satisfy $E(Y_{ij})=\mu$ and

$$\operatorname{cov}(Y_{ij},Y_{il})=\begin{cases}\sigma_\beta^2+\sigma_\varepsilon^2 & \text{if } j=l,\\ \sigma_\beta^2 & \text{otherwise,}\end{cases} \tag{2}$$

which is known as the homogeneous or exchangeable correlation structure.

The homogeneous correlation structure can also be obtained under the *transformation model*. Put

$$V=\sigma_\varepsilon^2 I_m+\sigma_\beta^2 J_m,$$

where I_m is the $m\times m$ identity matrix and $J_m=\mathbf{1}_m\mathbf{1}_m^{\mathrm{T}}$ is the $m\times m$ matrix of 1s ($\mathbf{1}_m$ is the m-vector of 1s), and let C denote the $mg\times mg$ block diagonal matrix with g blocks of V along the diagonal so $C=\operatorname{blockdiag}(V,\ldots,V)$. Let z be an mg-vector of independent and identically distributed random variables with mean 0, variance 1 and fourth moment ζ_4. Then let

$$y=\mathbf{1}_{mg}\mu+C^{1/2}z \tag{3}$$

so that

$$\begin{aligned}E(y)&=\mathbf{1}_{mg}\mu,\\ \operatorname{var}(y)&=C,\end{aligned} \tag{4}$$

corresponding to the mean and covariance structure (2). We use the symmetric matrix square root throughout.

Although models (1) and (3) produce the same correlation structure, they do not necessarily produce the same higher moment structure. For example, under both models, the within-block means $\bar{Y}_{i\cdot}$ (which determine the properties of the between sum of squares) are independent random variables with mean μ and variance $(\sigma_\varepsilon^2+m\sigma_\beta^2)/m$. However,

$$E\{(\bar{Y}_{i\cdot}-\mu)^4\}=\begin{cases}m^{-3}\{m^3E(\beta^4)+6m^2\sigma_\beta^2\sigma_\varepsilon^2+E(\varepsilon^4)+3(m-1)\sigma_\varepsilon^4\} & \text{random-effect model,}\\ m^{-3}(\sigma_\varepsilon^2+m\sigma_\beta^2)^2(\zeta_4-3+3m) & \text{transformation model.}\end{cases}$$

These fourth-order moments are the same under normality but are difficult to compare in general without specifying all the quantities that are involved in the expressions. In the particular case that β/σ_β and $\varepsilon/\sigma_\varepsilon$ have the same distribution as z, we only have to specify $(m,\gamma=\sigma_\beta/\sigma_\varepsilon,\zeta_4)$. For the Student t-distribution with 5 degrees of freedom, $\zeta_4=9$ and we obtain

$$E\{(\bar{Y}_{i\cdot}-\mu)^4\}=\begin{cases}3m^{-3}\sigma_\varepsilon^4(3m^3\gamma^4+2m^2\gamma^2+m+2) & \text{random-effect model,}\\ 3m^{-3}\sigma_\varepsilon^4(1+m\gamma^2)^2(m+2) & \text{transformation model.}\end{cases}$$

Fig. 1 partitions the (m,γ)-space for the Student t-distribution with 5 degrees of freedom into the region in which the within-group mean has greater fourth moment under the transformation model and the region in which it has greater fourth moment under the random-effect model. We see that, for this distribution, unless γ and m are small, the within-group mean has greater fourth moment under the random-effect than the transformation model. Although it is convenient for making comparisons, we should keep in mind that there is no reason in

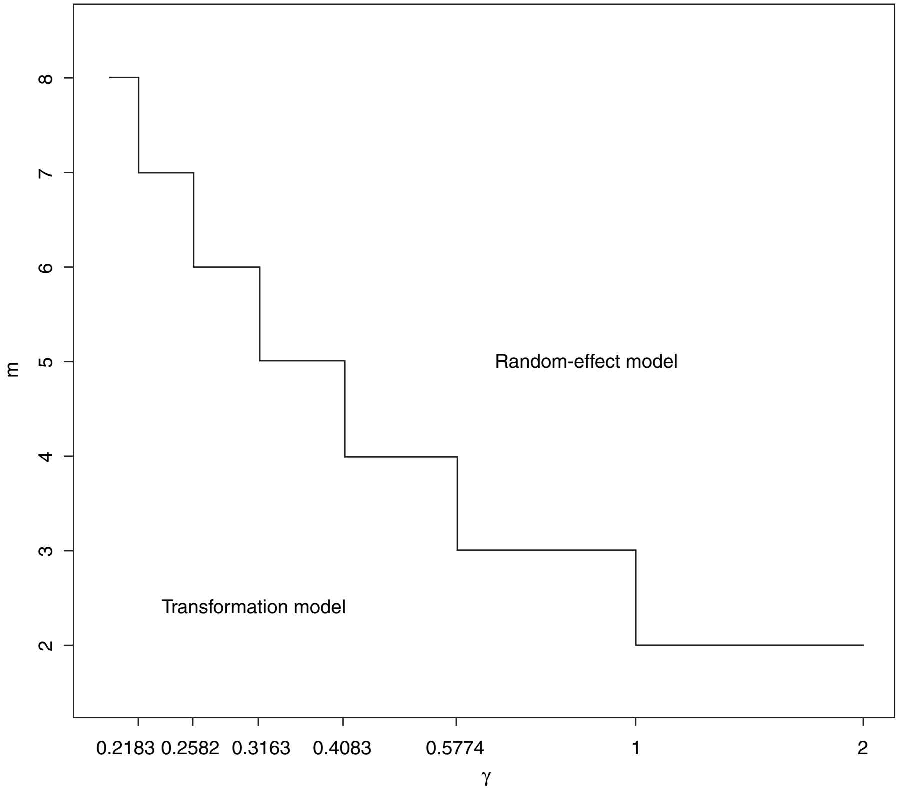

Fig. 1. Partition of the (m, γ) space for the Student t-distribution with 5 degrees of freedom into the region in which the within-group mean has greater fourth moment under the transformation model and the region in which it has greater fourth moment under the random-effect model: the γ-axis is on the $\log_2$-scale

practice to constrain β/σ_β, $\varepsilon/\sigma_\varepsilon$ and z to have the same distribution, never mind this particular distribution.

Our evaluation of various bootstrap resampling methods will be based on the first two moments of T, S_{B2} and S_{W2}. These moments are readily available under normality where (for both models) $E(T)=mg\mu$, $\text{var}(T)=mg(m\sigma_\beta^2+\sigma_\varepsilon^2)$, $E(S_{B2})=(g-1)(m\sigma_\beta^2+\sigma_\varepsilon^2)$, $\text{var}(S_{B2})=2(g-1)(m\sigma_\beta^2+\sigma_\varepsilon^2)^2$, $ES_{W2}=g(m-1)\sigma_\varepsilon^2$, $\text{var}(S_{W2})=2g(m-1)\sigma_\varepsilon^4$ and $\text{cov}(S_{B2}, S_{W2})=0$. However, we require the moments for non-normal distributions. In this case, we obtain different results for the two models.

Theorem 1. Suppose that the random-effect model (1) holds. Then the total T has mean $mg\mu$ and variance $mg(m\sigma_\beta^2+\sigma_\varepsilon^2)$. The sums of squares S_{B2} and S_{W2} have means $(g-1)(m\sigma_\beta^2+\sigma_\varepsilon^2)$ and $g(m-1)\sigma_\varepsilon^2$ respectively and

$$\text{var}(S_{B2})=\frac{g-1}{g}\left\{m^2(g-1)\,E(\beta^4)+4mg\sigma_\beta^2\sigma_\varepsilon^2+\frac{g-1}{m}E(\varepsilon^4)+\frac{2mg-3g+3}{m}\sigma_\varepsilon^4-m^2(g-3)\sigma_\beta^4\right\},$$

$$\text{var}(S_{W2})=\frac{(m-1)g}{m}\{(m-1)\,E(\varepsilon^4)-(m-3)\sigma_\varepsilon^4\},$$

and

$$\operatorname{cov}(S_{B2}, S_{W2}) = \frac{(m-1)(g-1)}{m}\{E(\varepsilon^4) - 3\sigma_\varepsilon^4\}. \tag{5}$$

The kinds of calculations that are used in the proof can be found in Tukey (1956) or Scheffé (1959), pages 343–346.

It is generally too stringent to require the bootstrap to reproduce the first two moments of a statistic in finite samples so we often impose the weaker requirement that this be achieved asymptotically. Corollary 1 is given for later reference. As discussed in Section 1, various asymptotic schemes are possible and we consider the two cases when $g \to \infty$ (with m fixed) and $g, m \to \infty$. We use the notation $a_n \sim b_n$ to mean that $a_n/b_n \to 1$ as $n \to \infty$ and in this case say that a_n is asymptotic to b_n.

Corollary 1. Suppose that the random-effect model (1) holds. Then, as $g \to \infty$ with m fixed, the mean of the total T is asymptotic to $mg\mu$ and the variance is asymptotic to $mg(m\sigma_\beta^2 + \sigma_\varepsilon^2)$. The means of the sums of squares S_{B2} and S_{W2} are asymptotic to $g(m\sigma_\beta^2 + \sigma_\varepsilon^2)$ and $(m-1)g\sigma_\varepsilon^2$ respectively and

$$\operatorname{var}(S_{B2}) \sim m^2 g\, E(\beta^4) + 4mg\sigma_\beta^2\sigma_\varepsilon^2 + \frac{g}{m} E(\varepsilon^4) + \frac{(2m-3)g}{m}\sigma_\varepsilon^4 - m^2 g\sigma_\beta^4,$$

$$\operatorname{var}(S_{W2}) \sim \frac{(m-1)g}{m}\{(m-1)\, E(\varepsilon^4) - (m-3)\sigma_\varepsilon^4\}$$

and

$$\operatorname{cov}(S_{B2}, S_{W2}) \sim \frac{(m-1)g}{m}\{E(\varepsilon^4) - 3\sigma_\varepsilon^4\}.$$

As $m,g \to \infty$, the mean of the total T is asymptotic to $mg\mu$ and the variance to $m^2 g\sigma_\beta^2$. The means of the sums of squares S_{B2} and S_{W2} are asymptotic to $mg\sigma_\beta^2$ and $mg\sigma_\varepsilon^2$ respectively and

$$\operatorname{var}(S_{B2}) \sim m^2 g\{E(\beta^4) - \sigma_\beta^4\},$$

$$\operatorname{var}(S_{W2}) \sim mg\{E(\varepsilon^4) - \sigma_\varepsilon^4\}$$

and

$$\operatorname{cov}(S_{B2}, S_{W2}) \sim g\{E(\varepsilon^4) - 3\sigma_\varepsilon^4\}.$$

Under the transformation model, the moments of the sums of squares are given in theorem 2.

Theorem 2. Suppose that the transformation model (3) holds. Then the sample total T has the same mean and variance as under the random-effect model, the sums of squares have the same means as under the random-effect model but

$$\operatorname{var}(S_{B2}) = \frac{g-1}{mg}(\sigma_\varepsilon^2 + m\sigma_\beta^2)^2\{(g-1)\zeta_4 + 2mg - 3g + 3\},$$

$$\operatorname{var}(S_{W2}) = \frac{(m-1)g}{m}\sigma_\varepsilon^4\{(m-1)\zeta_4 - (m-3)\}$$

and

$$\operatorname{cov}(S_{B2}, S_{W2}) = \frac{(m-1)(g-1)}{m}(\sigma_\varepsilon^2 + m\sigma_\beta^2)\sigma_\varepsilon^2(\zeta_4 - 3).$$

Proof. Write the sums of squares as quadratic forms and then use the fact that

$$E(z^{\mathrm{T}} Fzz^{\mathrm{T}} Gz) = \zeta_4 \sum_i F_{ii}G_{ii} + \sum_{i \neq j}(F_{ii}G_{jj} + F_{ij}G_{ij} + F_{ij}G_{ji}). \qquad \square$$

The variances and covariance of the sums of squares are different under the two models. In the special case that β/σ_β and $\varepsilon/\sigma_\varepsilon$ have the same distribution as z, the expressions are simpler to compare and we note the following properties.

(a) The expressions for the variance of the between sum of squares both follow from the fact that

$$\text{var}(S_{\text{B2}}) = \frac{m^2(g-1)^2}{g} E\{(\bar{Y}_{i\cdot}-\mu)^4\} - \frac{m^2(g-1)(g-3)}{g} E\{(\bar{Y}_{i\cdot}-\mu)^2\}^2.$$

Since $E(\bar{Y}_{i\cdot}-\mu)^2$ is the same under the two models, we can compare the variances of the between sum of squares by comparing $E\{(\bar{Y}_{i\cdot}-\mu)^4\}$ under both models as was done for the Student t-distribution with 5 degrees of freedom in Fig. 1.

(b) The variance of the within sum of squares is the same for both models.

(c) The covariance of the sums of squares is greater in absolute value for the transformation model than for the random-effect model.

As we noted before, there is no reason to constrain β/σ_β and $\varepsilon/\sigma_\varepsilon$ to have the same distribution. Some more general comparisons of the variances that were obtained in theorems 1 and 2 (not

Table 1. Comparison of the target values $(g-1)(m\sigma_\beta^2+\sigma_\varepsilon^2)$ and $g(m-1)\sigma_\varepsilon^2$ estimated by the between and within sums of squares respectively and the exact variances of the total, between and within sums of squares and the exact covariance of the between and within sums of squares under the random-effect model (theorem 1) and under the transformation model (theorem 2)†

g = *5*					
Target values	Between	20.00	22.67	30.67	33.33
	Within	15.00	25.00	15.00	25.00
Random effect	β–ε	*N*–*N*	*N*–*t*(5)	*t*(5)–*N*	*t*(5)–*t*(5)
	Total	100.00	113.33	153.33	166.67
	Between	200.00	270.22	1323.56	1422.22
	Within	30.00	270.83	30.00	270.83
	Covariance of sums of squares	0.00	50.00	0.00	50.00
Transformation	z, $(\sigma_\beta^2, \sigma_\varepsilon^2)$	*N* (1,1)	*t**(5) (1,5/3)	*t**(5) (5/3,1)	*t**(5) (5/3,5/3)
	Total	100.00	113.33	153.33	166.67
	Between	200.00	411.02	752.36	888.89
	Within	30.00	270.83	97.50	270.83
	Covariance of sums of squares	0.00	170.00	138.00	250.00
g = *15*					
Target values	Between	70.00	79.33	107.33	116.67
	Within	45.00	75.00	45.00	75.00
Random effect	β–ε	*N*–*N*	*N*–*t*(5)	*t*(5)–*N*	*t*(5)–*t*(5)
	Total	300.00	340.00	460.00	500.00
	Between	700.00	953.56	5130.22	5483.33
	Within	90.00	812.50	90.00	812.50
	Covariance of sums of squares	0.00	175.00	0.00	175.00
Transformation	z, $(\sigma_\beta^2, \sigma_\varepsilon^2)$	*N* (1,1)	*t**(5) (1,5/3)	*t**(5) (5/3,1)	*t**(5) (5/3,5/3)
	Total	300.00	340.00	460.00	500.00
	Between	700.00	1528.49	2797.82	3305.56
	Within	90.00	812.50	292.50	812.50
	Covariance of sums of squares	0.00	595.00	483.00	875.00

†The results are presented for $m=4$ and $g=5$ or $g=15$. The symbol N denotes the standard normal distribution, $t(5)$ the Student t-distribution with 5 degrees of freedom and $t^*(5)=\sqrt{(3/5)}t(5)$ the standardized Student t-distribution with 5 degrees of freedom. For the random-effect model, the distributions are given as that of the random effect β followed by that of ε; for the transformation model, the distribution is that of z followed by the values of $(\sigma_\beta^2, \sigma_\varepsilon^2)$. The settings have been chosen so that the target values are the same under both models.

necessarily assuming that the distributions are the same) are given in Table 1. We can choose a wide variety of settings for the random-effect and transformation models. For consistency we have arranged for the between and within sums of squares to estimate the same target value under the two models (although the targets change with g and the particular distributions). As we would expect, the variance of the sample total is the same under the two models because it depends only on the second-moment structures, which agree under the two models. Although the variances and covariance of the sums of squares are the same for the two models under normality, in general, the variance of the between and within sum of squares can be substantially different under the two models. The covariance is the same as or larger in the transformation model than in the random-effect model. Note that the first (the $N-N$ or N (1,1)) and the last (the $t(5)$–$t(5)$ or $t^*(5)$ (5/3, 5/3)) columns in Table 1 are covered by Fig. 1 (and the numbers are consistent with the conclusions that were drawn above) but the other two columns are not. It is interesting that the size of the differences in the variances under the two models can be so large.

3. Bootstrap methods

We now consider the issue of bootstrapping to obtain estimates of the variance and covariance of the between sum of squares S_{B2} and the within sum of squares S_{W2} along with results for the sample total T. There are various seemingly attractive approaches that can be used to obtain bootstrap estimates and we explore several of them.

For each bootstrap, we give the exact bootstrap means and variances of the total, between and within sums of squares. Since these bootstrap calculations are made conditionally on the observed data, they depend on the type of bootstrap but not on the underlying model. We then describe the asymptotic behaviour of these bootstrap moments which depends on the convergence of sample statistics to population quantities of the underlying model. Hence the convergence results do depend on the underlying model.

3.1. Random-effect bootstrap

The random-effect bootstrap method is based on the random-effect model: we construct predictors $\hat{\beta}_i$ and $\hat{\varepsilon}_{ij}$ of β_i and ε_{ij} respectively which satisfy $\bar{\hat{\beta}}=0$ and $\bar{\hat{\varepsilon}}=0$, draw a sample of size g independently with replacement from $\{\hat{\beta}_i : i=1,\ldots,g\}$ to produce the bootstrap sample $\{\hat{\beta}_i^* : i=1,\ldots,g\}$ and then independently draw a sample of size mg independently with replacement from $\{\hat{\varepsilon}_{ij} : j=1,\ldots,m, i=1,\ldots,g\}$ to produce the bootstrap sample $\{\hat{\varepsilon}_{ij}^* : j=1,\ldots,m,$ $i=1,\ldots,g\}$. The bootstrap observations are constructed as suggested by the random-effect model (1) as

$$Y_{ij}^* = \bar{Y}_{..} + \hat{\beta}_i^* + \hat{\varepsilon}_{ij}^*, \qquad j=1,\ldots,m, \quad i=1,\ldots,g.$$

The $\hat{\varepsilon}_{ij}^*$ are independent and identically distributed random variables with bootstrap mean $\bar{\hat{\varepsilon}}=0$, variance

$$u_{\varepsilon 2} = (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} \hat{\varepsilon}_{ij}^2$$

and fourth central moment

$$u_{\varepsilon 4} = (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} \hat{\varepsilon}_{ij}^4.$$

They are independent of the $\hat{\beta}_i^*$ which are independent and identically distributed random variables with bootstrap mean $\bar{\hat{\beta}}=0$, variance

$$u_{\beta 2}=g^{-1}\sum_{i=1}^{g}\hat{\beta}_i^2$$

and fourth moment

$$u_{\beta 4}=g^{-1}\sum_{i=1}^{g}\hat{\beta}_i^4.$$

It follows that the observations Y_{ij}^* and Y_{kl}^* are (conditionally) independent when $i\neq k$, have $E^*(Y_{ij}^*)=\bar{Y}_{..}$ and

$$\operatorname{cov}^*(Y_{ij}^*,Y_{il}^*)=\begin{cases}u_{\beta 2}+u_{\varepsilon 2} & \text{if } j=l,\\ u_{\beta 2} & \text{otherwise.}\end{cases}$$

Comparing this result with expression (2), we see that the bootstrap observations have the same first- and second-moment structure as the original data. Also,

$$E^*\{(\bar{Y}_{i.}^*-\bar{Y}_{..})^4\}=u_{\beta 4}+6u_{\beta 2}u_{\varepsilon 2}/m+\{u_{\varepsilon 4}+3(m-1)u_{\varepsilon 2}^2\}/m^3,$$

which has the same form as the fourth moment of the within-group mean under the random-effect model. The expected bootstrap moments depend on the choice of the predictors $\hat{\beta}_i$ and $\hat{\varepsilon}_{ij}$.

The properties of the random-effect bootstrap statistics follow from theorem 1.

Theorem 3. The random-effects bootstrap mean and variance of T^* are T and $m^2gu_{\beta 2}+mgu_{\varepsilon 2}$ respectively. The random-effect bootstrap means of $S_{\mathrm{B}2}^*$ and $S_{\mathrm{W}2}^*$ are $(g-1)(mu_{\beta 2}+u_{\varepsilon 2})$ and $(m-1)gu_{\varepsilon 2}$ respectively. The random-effect bootstrap variances and covariances are

$$\operatorname{var}^*(S_{\mathrm{B}2}^*)=\frac{g-1}{g}\left\{m^2(g-1)u_{\beta 4}+4mgu_{\beta 2}u_{\varepsilon 2}+\frac{g-1}{m}u_{\varepsilon 4}+\frac{2mg-3g+3}{m}u_{\varepsilon 2}^2-m^2(g-3)u_{\beta 2}^2\right\},$$

$$\operatorname{var}^*(S_{\mathrm{W}2}^*)=\frac{g(m-1)^2}{m}u_{\varepsilon 4}-\frac{g(m-1)(m-3)}{m}u_{\varepsilon 2}^2$$

and

$$\operatorname{cov}^*(S_{\mathrm{B}2}^*,S_{\mathrm{W}2}^*)=\frac{(m-1)(g-1)}{m}(u_{\varepsilon 4}-3u_{\varepsilon 2}^2).$$

It follows that the random-effect bootstrap variance of the total and the random-effect bootstrap estimates of the sums of squares are consistent under the random-effect model provided that $u_{\beta 2}\sim\sigma_\beta^2$ and $u_{\varepsilon 2}\sim\sigma_\varepsilon^2$. Since the expressions for the variances and covariance of the sums of squares in theorem 3 are of the same form as those in theorem 1, it follows that, under the random-effect model, the bootstrap variances of the sums of squares are consistent provided, in addition, that $u_{\beta 4}\sim E(\beta^4)$ and $u_{\varepsilon 4}\sim E(\varepsilon^4)$. They are not consistent under the transformation model.

The usual predictor of ε_{ij} is $\tilde{\varepsilon}_{ij}=Y_{ij}-\bar{Y}_{i.}$ and the estimated best linear unbiased predictor of β_i (when the variance components are known) is

$$\tilde{\beta}_i=\frac{S_{\mathrm{B}2}/(g-1)-S_{\mathrm{W}2}/g(m-1)}{S_{\mathrm{B}2}/g}(\bar{Y}_{i.}-\bar{Y}_{..}),$$

provided that $S_{\mathrm{B}2}/(g-1)>S_{\mathrm{W}2}/g(m-1)$ (Searle *et al.* (1992), page 55). Note that, asymptotically, we can estimate the denominator just as well by $S_{\mathrm{B}2}/(g-1)$, which matches the numerator better. However, as we shall see below, the choice of g rather than $g-1$ leads to smaller bias. The sum of squares $g^{-1}\Sigma_{i=1}^{g}\tilde{\beta}_i^2$ converges to $(m\sigma_\beta^2+\sigma_\varepsilon^2)^{-1}m\sigma_\beta^4$, which is too small. We can obtain the result required if we set

$$\hat{\beta}_i=\sqrt{\left\{\frac{S_{\mathrm{B}2}/(g-1)-S_{\mathrm{W}2}/g(m-1)}{S_{\mathrm{B}2}/g}\right\}}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})$$

so that

$$u_{\beta 2}=\frac{1}{g}\sum_{i=1}^{g}\hat{\beta}_i^2=\frac{1}{m(g-1)}S_{\mathrm{B}2}-\frac{1}{m(m-1)g}S_{\mathrm{W}2}$$

is the usual unbiased (and consistent) estimator of σ_β^2. The predictor $\hat{\beta}_i$ is still shrunk towards 0 but not as much as $\tilde{\beta}_i$. Similarly, we define the expanded predictor

$$\hat{\varepsilon}_{ij}=\sqrt{\left(\frac{m}{m-1}\right)}(Y_{ij}-\bar{Y}_{i\cdot})$$

so that

$$u_{\varepsilon 2}=\frac{1}{mg}\sum_{i=1}^{g}\sum_{j=1}^{m}\hat{\varepsilon}_{ij}^2=\frac{1}{g(m-1)}S_{\mathrm{W}2}$$

is unbiased for σ_ε^2. These are the same as the adjustments that were recommended by Davison and Hinkley (1997), page 102. With this choice of predictors,

$$\operatorname{cov}^*(Y_{ij}^*,Y_{il}^*)=\begin{cases}\dfrac{1}{m(g-1)}S_{\mathrm{B}2}+\dfrac{1}{mg}S_{\mathrm{W}2} & \text{if } j=l,\\ \dfrac{1}{m(g-1)}S_{\mathrm{B}2}-\dfrac{1}{m(m-1)g}S_{\mathrm{W}2} & \text{otherwise.}\end{cases}$$

This looks more like the covariance structure (2) if, for $j\neq l$, we write $\operatorname{cov}^*(Y_{ij}^*,Y_{il}^*)=\operatorname{var}^*(Y_{ij}^*)-\{(m-1)g\}^{-1}S_{\mathrm{W}2}$ and the expected bootstrap moments equal expression (2) exactly.

For the fourth moments,

$$\begin{aligned}u_{\beta 4}&=\left\{\frac{S_{\mathrm{B}2}/(g-1)-S_{\mathrm{W}2}/g(m-1)}{S_{\mathrm{B}2}/g}\right\}^2\frac{1}{g}\sum_{i=1}^{g}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^4\\&=\left\{\frac{S_{\mathrm{B}2}/(g-1)-S_{\mathrm{W}2}/g(m-1)}{S_{\mathrm{B}2}/g}\right\}^2\frac{1}{mg}S_{\mathrm{B}4}\end{aligned}$$

and

$$\begin{aligned}u_{\varepsilon 4}&=\frac{m}{(m-1)^2g}\sum_{i=1}^{g}\sum_{j=1}^{m}(Y_{ij}-\bar{Y}_{i\cdot})^4\\&=\frac{m}{(m-1)^2g}S_{\mathrm{W}4}.\end{aligned}$$

The asymptotic version of theorem 3 follows from the fact that as $g\rightarrow\infty$ with m fixed

$$\frac{1}{mg}S_{\mathrm{B}4}\sim E\{(\beta_i+\bar{\varepsilon}_{i\cdot})^4\}=\frac{1}{m^3}\{m^3E(\beta^4)+6m^2\sigma_\beta^2\sigma_\varepsilon^2+E(\varepsilon^4)+3(m-1)\sigma_\varepsilon^4\}$$

and

$$\frac{1}{mg} S_{\mathrm{W4}} \sim \frac{1}{m^3}\{(m^3-4m^2+6m-3)\, E(\varepsilon^4)+3(m-1)(2m-1)\sigma_\varepsilon^4\}.$$

Corollary 2. Suppose that the random-effect model holds. As $g \to \infty$ with m fixed, the random-effect bootstrap mean and variance of T^* are $mg\mu$ and $mg(m\sigma_\beta^2+\sigma_\varepsilon^2)$ respectively. The random-effect bootstrap means of S_{B2}^* and S_{W2}^* are $g(m\sigma_\beta^2+\sigma_\varepsilon^2)$ and $g(m-1)\sigma_\varepsilon^2$ respectively. The random-effect bootstrap variances and covariances are

$$\begin{aligned}\operatorname{var}^*(S_{\mathrm{B2}}^*) \sim{}& m^4 g \frac{\sigma_\beta^4}{(m\sigma_\beta^2+\sigma_\varepsilon^2)^2} E\{(\beta_i+\bar\varepsilon_{i\cdot})^4\}+4mg\sigma_\beta^2\sigma_\varepsilon^2+\frac{g}{m^2(m-1)^2}\{(m^3-4m^2+6m-3)\, E(\varepsilon^4)\\ &+3(m-1)(2m-3)\sigma_\varepsilon^4\}+\frac{(2m-3)g}{m}\sigma_\varepsilon^4-m^2 g\sigma_\beta^4,\\ \operatorname{var}^*(S_{\mathrm{W2}}^*) \sim{}& \frac{g}{m^2}\{(m^3-4m^2+6m-3)\, E(\varepsilon^4)+3(m-1)(2m-3)\sigma_\varepsilon^4\}-\frac{(m-3)(m-1)g}{m}\sigma_\varepsilon^4\end{aligned}$$

and

$$\operatorname{cov}^*(S_{\mathrm{B2}}^*, S_{\mathrm{W2}}^*) \sim \frac{g}{m^2(m-1)}\{(m^3-4m^2+6m-3)E(\varepsilon^4)+3(m-1)(2m-3)\sigma_\varepsilon^4-3m(m-1)^2\sigma_\varepsilon^4\}.$$

As $m, g \to \infty$, the random-effect bootstrap variances of T^*, S_{B2}^* and S_{W2}^* and the covariances between the sums of squares are asymptotically correct.

As we have seen, the bootstrap observations have the same first-, second- and fourth-moment structure as the original data under the random-effect model. The bootstrap estimates the first and second moments of the total and the first moments of the between and within sums of squares as $g \to \infty$ with m fixed. However, the bootstrap estimates the variance and covariance of the between and within sums of squares when $m, g \to \infty$ but not when $g \to \infty$ with m fixed.

3.2. Residual bootstrap

The residual bootstrap is motivated by the transformation model: the idea is to compute residuals which can be bootstrapped and then used to reconstruct bootstrap observations.

Let $\hat V$ and $\hat C$ denote the matrices V and C respectively, with the unknown variances σ_ε^2 and σ_β^2 replaced by estimators. Then define the residuals

$$r = \hat C^{-1/2}(y - \mathbf{1}_{mg}\bar Y_{\cdot\cdot})$$

and the centred and standardized residuals

$$r_{\mathrm{c}} = \{r - (mg)^{-1}\mathbf{1}_{mg}\mathbf{1}_{mg}^{\mathrm{T}} r\}/s_r,$$

where $s_r^2 = (mg)^{-1}\{r^{\mathrm{T}} r - (mg)^{-1}(\mathbf{1}_{mg}^{\mathrm{T}} r)^2\}$. We obtain a bootstrap sample by drawing a sample r_{c}^* of size mg independently with replacement from r_{c} and setting

$$y^* = \mathbf{1}_{nb}\bar Y_{\cdot\cdot} + \hat C^{1/2} r_{\mathrm{c}}^*.$$

The residuals have bootstrap mean 0, variance I_{mg} and

$$E^*\{(r_{\mathrm{c}}^* \otimes r_{\mathrm{c}}^*)(r_{\mathrm{c}}^* \otimes r_{\mathrm{c}}^*)^{\mathrm{T}}\} = E^*(r_{\mathrm{c}i}^* r_{\mathrm{c}j}^* r_{\mathrm{c}k}^* r_{\mathrm{c}l}^*),$$

where

$$E^*(r_{\mathrm{c}i}^* r_{\mathrm{c}j}^* r_{\mathrm{c}k}^* r_{\mathrm{c}l}^*) = \begin{cases} (mg)^{-1}\sum_{i=1}^{mg} r_{ci}^4 & \text{if } i=j=k=l,\\ 1 & \text{if } i=j\neq k=l, i=k\neq j=l, \text{ or } i=l\neq j=k,\\ 0 & \text{otherwise.}\end{cases}$$

It follows that $E^*(y^*)=\mathbf{1}_{mg}\bar{Y}_{..}$ and $\mathrm{var}^*(y^*)=\hat{C}$, which is an empirical version of expression (4) so the bootstrap sample y^* has the same first- and second-moment structure as the original data y. The expected bootstrap moments are $E\{E^*(y^*)\}=\mathbf{1}_{nb}\mu$ and $E\{\mathrm{var}^*(y^*)\}=E(\hat{C})$ so the first two moments match exactly if $E(\hat{C})=C$. For the within-group mean, we have

$$E^*\{(\bar{Y}^*_{i\cdot}-\bar{Y}^*_{..})^4\}=m^{-3}(\hat{\sigma}^2_\varepsilon+m\hat{\sigma}^2_\beta)^2(\hat{\zeta}_4-3+3m),$$

where $\hat{\sigma}_\beta$ and $\hat{\sigma}_\varepsilon$ are the estimates of σ_β and σ_ε that are used in $\hat{C}$, and

$$\hat{\zeta}_4=(mg)^{-1}\sum_{i=1}^{mg} r^4_{\mathrm{c}i}.$$

We show in the proof of corollary 3 below that this moment is consistent under the transformation model; it is not consistent under the random-effect model.

Theorem 4. The residual bootstrap mean and variance of T^* are T and $mg(\hat{\sigma}^2_\varepsilon+m\hat{\sigma}^2_\beta)$ respectively. The residual bootstrap means of $S^*_{\mathrm{B}2}$ and $S^*_{\mathrm{W}2}$ are $(g-1)(\hat{\sigma}^2_\varepsilon+m\hat{\sigma}^2_\beta)$ and $g(m-1)\hat{\sigma}^2_\varepsilon$ respectively. The residual bootstrap variances and covariances are

$$\mathrm{var}(S_{\mathrm{B}2})=\frac{g-1}{mg}(\hat{\sigma}^2_\varepsilon+m\hat{\sigma}^2_\beta)^2\{(g-1)\hat{\zeta}_4+2mg-3g+3\},$$

$$\mathrm{var}(S_{\mathrm{W}2})=\frac{(m-1)g}{m}\hat{\sigma}^4_\varepsilon\{(m-1)\hat{\zeta}_4+m-3\}$$

and

$$\mathrm{cov}(S_{\mathrm{B}2},S_{\mathrm{W}2})=\frac{(m-1)(g-1)}{m}(\hat{\sigma}^2_\varepsilon+m\hat{\sigma}^2_\beta)\hat{\sigma}^2_\varepsilon(\hat{\zeta}_4-3).$$

The proof is very similar to that for theorem 2 and is omitted.

Theorem 4 shows that the residual bootstrap variances and covariance are of the same form as the variances and covariance that are given in theorem 2 with $\hat{C}$ in place of C and r_{c} in place of z. As a result, the consistency of the bootstrap variance estimates under the transformation model becomes a question of the consistency of the estimates, $\hat{C}$ and the second and fourth moments of the r_cs. We have the following result.

Corollary 3. Suppose that the transformation model holds. Provided that $\hat{V}\sim V$ as $g\to\infty$ with m fixed, the residual bootstrap means and variances of T^*, $S^*_{\mathrm{B}2}$ and $S^*_{\mathrm{W}2}$ and the bootstrap covariance between $S^*_{\mathrm{B}2}$ and $S^*_{\mathrm{W}2}$ are asymptotically correct as $g\to\infty$ with m fixed.

Proof. Since $\hat{V}\sim V$, the result follows if we can show that $s^2_r=1+o_p(1)$ and

$$(mg)^{-1}\sum_{i=1}^{mg} r^4_{\mathrm{c}i}=\zeta_4+o_p(1).$$

We have

$$\begin{aligned}\bar{r}&=\frac{1}{mg}\mathbf{1}^{\mathrm{T}}_{mg}\hat{C}^{-1/2}(y-\mathbf{1}_{mg}\bar{Y}_{..})\\&=\frac{1}{mg}\mathbf{1}^{\mathrm{T}}_{m}\hat{V}^{-1/2}\sum_{i=1}^{g}(y_i-\mathbf{1}_m\bar{Y}_{..})\\&=O_p(g^{-1/2})\end{aligned}$$

so

$$
\begin{aligned}
s_r^2 &= \frac{1}{mg} r^{\mathrm{T}} r - \bar{r}^2 \\
&= \frac{1}{mg} (y - \mathbf{1}_{mg} \bar{Y}_{..})^{\mathrm{T}} \hat{C}^{-1} (y - \mathbf{1}_{mg} \bar{Y}_{..}) + O_p(g^{-1}) \\
&= \frac{1}{mg} \sum_{i=1}^{g} (y_i - \mathbf{1}_m \bar{Y}_{..})^{\mathrm{T}} \hat{V}^{-1} (y_i - \mathbf{1}_m \bar{Y}_{..}) + O_p(g^{-1}) \\
&= \mathrm{tr}\left\{ \hat{V}^{-1} \frac{1}{mg} \sum_{i=1}^{g} (y_i - \mathbf{1}_m \bar{Y}_{..})(y_i - \mathbf{1}_m \bar{Y}_{..})^{\mathrm{T}} \right\} + O_p(g^{-1}) \\
&= \frac{1}{m} \mathrm{tr}(I_m) + o_p(1) + O_p(g^{-1}).
\end{aligned}
$$

Next,

$$
\begin{aligned}
(mg)^{-1} \sum_{i=1}^{mg} r_{ci}^4 &= (mg)^{-1} \sum_{i=1}^{mg} (r_i - \bar{r})^4 / s_r^4 \\
&= \left\{ (mg)^{-1} \sum_{i=1}^{mg} r_i^4 - 4(mg)^{-1} \sum_{i=1}^{mg} r_i^3 \bar{r} + 6(mg)^{-1} \sum_{i=1}^{mg} r_i^2 \bar{r}^2 - 3\bar{r}^4 \right\} \Big/ s_r^4
\end{aligned}
$$

so we just need to establish that

$$
(mg)^{-1} \sum_{i=1}^{mg} r_i^4 = \zeta_4 + o_p(1).
$$

Let δ_j be the n-vector with jth component equal to 1 and all other components 0. Note that, in the following expression, we can assume that $\mu = 0$. Then we have

$$
\begin{aligned}
(mg)^{-1} \sum_{i=1}^{mg} r_i^4 &= (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} \{\delta_j^{\mathrm{T}} \hat{V}^{-1/2} (y_i - \bar{Y}_{..})\}^4 \\
&= (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} y_i)^4 - 4(mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} y_i)^3 \delta_j^{\mathrm{T}} \hat{V}^{-1/2} \bar{Y}_{..} \\
&\quad + 6(mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} y_i)^2 (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} \bar{Y}_{..})^2 - 3m^{-1} \sum_{j=1}^{m} (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} \bar{Y}_{..})^4. \qquad (6)
\end{aligned}
$$

Write $z^{\mathrm{T}} = (z_1^{\mathrm{T}}, \ldots, z_g^{\mathrm{T}})$. Then, noting that $\mathrm{vec}(z_i z_i^{\mathrm{T}}) = z_i \otimes z_i$, we have

$$
\begin{aligned}
(mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} y_i)^4 &= (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} (\delta_j^{\mathrm{T}} \hat{V}^{-1/2} V^{1/2} z_i)^4 \\
&= (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} \{(V^{1/2} \hat{V}^{-1/2} \delta_j \otimes V^{1/2} \hat{V}^{-1/2} \delta_j)^{\mathrm{T}} (z_i \otimes z_i)\}^2 \\
&= (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} \{(V^{1/2} \hat{V}^{-1/2} \delta_j \otimes V^{1/2} \hat{V}^{-1/2} \delta_j)^{\mathrm{T}} (z_i \otimes z_i)\} \\
&\qquad \times \{(z_i \otimes z_i)^{\mathrm{T}} (V^{1/2} \hat{V}^{-1/2} \delta_j \otimes V^{1/2} \hat{V}^{-1/2} \delta_j)\} \\
&= (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} \mathrm{tr}\{(V^{1/2} \hat{V}^{-1/2} \delta_j \otimes V^{1/2} \hat{V}^{-1/2} \delta_j)(V^{1/2} \hat{V}^{-1/2} \delta_j \\
&\qquad \otimes V^{1/2} \hat{V}^{-1/2} \delta_j)^{\mathrm{T}} (z_i \otimes z_i)(z_i \otimes z_i)^{\mathrm{T}}\}
\end{aligned}
$$

$$
\begin{aligned}
&= m^{-1}\sum_{j=1}^{m} \mathrm{tr}\{(V^{1/2}\hat{V}^{-1/2}\delta_j \otimes V^{1/2}\hat{V}^{-1/2}\delta_j)(V^{1/2}\hat{V}^{-1/2}\delta_j \otimes V^{1/2}\hat{V}^{-1/2}\delta_j)^{\mathrm{T}} \\
&\quad \times g^{-1}\sum_{i=1}^{g}(z_i \otimes z_i)(z_i \otimes z_i)^{\mathrm{T}}\} \\
&= m^{-1}\sum_{j=1}^{m} \mathrm{tr}[(\delta_j \otimes \delta_j)(\delta_j \otimes \delta_j)^{\mathrm{T}}\, E\{(z_i \otimes z_i)(z_i \otimes z_i)^{\mathrm{T}}\}] + o_p(1) \\
&= m^{-1}\sum_{j=1}^{m} E[\{(\delta_j \otimes \delta_j)(z_i \otimes z_i)\}^2] + o_p(1) \\
&= m^{-1}\sum_{j=1}^{m} E\{(\delta_j^T z_i)^4\} + o_p(1) \\
&= \zeta_4 + o_p(1).
\end{aligned}
$$

The result follows from the fact that $\bar{Y}_{..} = O_p(g^{-1/2})$ so the remaining terms in equation (6) are all of smaller order than the leading fourth-power term. □

The Davison and Hinkley (1997) criterion requiring that the bootstrap observations have the same low order moment structure as the original data cannot distinguish between the residual and random-effect bootstraps.

3.3. Cluster bootstrap

The simple cluster bootstrap is based on drawing samples of g clusters independently with replacement. The bootstrap sample is the set of mg observations Y_{ij}^* where the g m-vectors $(Y_{i1}^*, \ldots, Y_{im}^*)$ corresponding to clusters are independently and identically distributed with the distribution placing probability $1/g$ on each of the g m-vectors $(Y_{i1}, \ldots, Y_{im})$. This means that the observations Y_{ij}^* and Y_{kl}^* are bootstrap independent whenever $i \neq k$ with

$$E^*(Y_{ij}^*) = g^{-1}\sum_{i=1}^{g} Y_{ij} = \bar{Y}_{.j}$$

and

$$\mathrm{cov}^*(Y_{ij}^*, Y_{il}^*) = \begin{cases} \dfrac{1}{g}\sum_{i=1}^{g}(Y_{ij} - \bar{Y}_{.j})^2 & \text{if } j = l, \\ \dfrac{1}{g}\sum_{i=1}^{g}(Y_{ij} - \bar{Y}_{.j})(Y_{il} - \bar{Y}_{.l}) & \text{otherwise.} \end{cases}$$

This is not the same first two moment structure as the original data shown in expression (2), although $E\{E^*(Y_{ij}^*)\} = \mu$, $E\{\mathrm{var}^*(Y_{ij}^*)\} = g^{-1}(g-1)(\sigma_\beta^2 + \sigma_\varepsilon^2)$ and $E\{\mathrm{cov}^*(Y_{ij}^*, Y_{il}^*)\} = g^{-1} \times (g-1)\sigma_\beta^2$, so this structure is recovered as $g \to \infty$. Alternatively, as in the random-effect bootstrap, we can rescale the observations by $\surd\{g/(g-1)\}$ to make the result exact. For the within-group mean, we find that

$$E^*\{(\bar{Y}_{i.}^* - \bar{Y}_{..})^4\} = \frac{1}{mg} S_{\mathrm{B}4} \sim E\{(\bar{Y}_{i.} - \mu)^4\}$$

which is correct under either model. This is the main advantage of using a non-parametric bootstrap which is not based on a particular model—it is possible to be consistent under different models.

The above moment calculations treat the cluster vectors as exchangeable observations but do not take into account the fact that the components of these vectors are also exchangeable. Following Davison and Hinkley (1997) and McCullagh (2000), we can do this by permuting the observations within the clusters. This produces the randomized cluster bootstrap. The permutation step has no effect on our statistics, which are all invariant to permutation, but does change the calculations for the moments of individual observations as these are not invariant to permutation. In particular,

$$E^*(Y_{ij}^*) = (mg)^{-1} \sum_{i=1}^{g} \sum_{j=1}^{m} Y_{ij} = \bar{Y}_{..}$$

and

$$\operatorname{cov}^*(Y_{ij}^*, Y_{il}^*) = \begin{cases} \dfrac{1}{mg}(S_{\mathrm{B}2} + S_{\mathrm{W}2}) & \text{if } j = l, \\ \dfrac{1}{mg}\{S_{\mathrm{B}2} - (m-1)^{-1} S_{\mathrm{W}2}\} & \text{otherwise.} \end{cases}$$

As with the random-effect bootstrap, this looks more like the covariance structure (2) if, for $j \neq l$, we write

$$\operatorname{cov}^*(Y_{ij}^*, Y_{il}^*) = \operatorname{var}^*(Y_{ij}^*) - \{(m-1)g\}^{-1} S_{\mathrm{W}2}$$

but note that here the expected variance

$$E\{\operatorname{var}^*(Y_{ij}^*)\} = (mg)^{-1}\{m(g-1)\sigma_\beta^2 + (mg-1)\sigma_\varepsilon^2\}$$

and the expected covariance

$$E\{\operatorname{cov}^*(Y_{ij}^*, Y_{il}^*)\} = (mg)^{-1}\{m(g-1)\sigma_\beta^2 - \sigma_\varepsilon^2\}$$

which are both too small and only correct when $g \to \infty$.

The first two cluster bootstrap (and randomized cluster bootstrap) moments of the total, between and within sums of squares are given in theorem 5.

Theorem 5. The cluster bootstrap mean and variance of T^* are T and $mg^2 S_{\mathrm{B}2}$ respectively. The cluster bootstrap means of $S_{\mathrm{B}2}^*$ and $S_{\mathrm{W}2}^*$ are $(g-1)g^{-1} S_{\mathrm{B}2}$ and $S_{\mathrm{W}2}$ respectively. The cluster bootstrap variances and covariances are

$$\operatorname{var}^*(S_{\mathrm{B}2}^*) = \frac{m(g-1)^2}{g^2} S_{B4} - \frac{(g-1)(g-3)}{g^3} S_{\mathrm{B}2}^2,$$

$$\operatorname{var}^*(S_{\mathrm{W}2}^*) = \sum_{i=1}^{g} S_{\mathrm{W}2i}^2 - \frac{1}{g} S_{\mathrm{W}2}^2$$

and

$$\operatorname{cov}^*(S_{\mathrm{B}2}^*, S_{\mathrm{W}2}^*) = \frac{m(g-1)}{g} \sum_{i=1}^{g} S_{\mathrm{W}2i}(\bar{Y}_{i\cdot} - \bar{Y}_{..})^2 - \frac{g-1}{g^2} S_{\mathrm{W}2} S_{\mathrm{B}2}$$

respectively.

Proof. From general results on simple random sampling with replacement (e.g. Särndal *et al.* (1992), page 52), the cluster bootstrap sample total T^* is unbiased for the 'population' total T and has variance nS_{B2}. The results for the cluster bootstrap between sum of squares S_{B2}^* follow from the fact that S_{B2}^* is the scaled sample variance of the independent random variables $\bar{Y}_{i.}^*$. For the cluster bootstrap within sum of squares, we obtain

$$E(S_{W2}^*) = \sum_{i=1}^{g} S_{W2i} = S_{W2}$$

and

$$\begin{aligned}\operatorname{var}^*(S_{W2}^*) &= \frac{g-1}{g} \sum_{i=1}^{g} S_{W2i}^2 - \frac{1}{g} \sum_{i \neq h}^{g} S_{W2i} S_{W2h} \\ &= \frac{g-1}{g} \sum_{i=1}^{g} S_{W2i}^2 - \frac{1}{g} \left(\sum_{i=1}^{g} S_{W2i} \right)^2 + \frac{1}{g} \sum_{i=1}^{g} S_{W2i}^2 \\ &= \sum_{i=1}^{g} S_{W2i}^2 - \frac{1}{g} S_{W2}^2.\end{aligned}$$

The proof of the expression for the covariance is given in Appendix A. □

As we noted, the cluster bootstrap has the advantage of being non-parametric and not depending on any model. This makes it possible for the bootstrap estimates of the means, variances and covariances to be consistent under both models.

Corollary 4. Under both the random-effect and the transformation model, the cluster bootstrap variances of T^*, S_{B2}^* and S_{W2}^* and the covariance between the sums of squares are asymptotically correct as $g \to \infty$ with m fixed.

Corollary 4 shows that the cluster bootstrap applied to groups is a sensible estimator of the sampling variances of both the mean and the variance components in large samples. The cluster bootstrap variance of the mean is exactly the variance of the mean if we scale the block means by $\sqrt{\{g/(g-1)\}}$. This is not surprising because, under both models, the group vectors are simply independent and identically distributed random vectors.

3.4. Other bootstraps

We have not yet considered the two-stage bootstrap and the reverse two-stage bootstrap that were described in Section 1. The results for the cluster bootstrap suggest that excess variation is generated in these bootstraps. McCullagh's (2000) results show that the two-stage and reverse two-stage bootstraps do not produce consistent estimators for the variance of T^* unless both m and g tend to ∞. Thus, the two-stage and reverse two-stage bootstraps are inferior to the other bootstraps. Under the two-stage bootstrap

$$\operatorname{cov}^*(Y_{ij}^*, Y_{il}^*) = \begin{cases} \dfrac{1}{mg}(S_{B2} + S_{W2}) & \text{if } j = l, \\ \dfrac{1}{mg} S_{B2} & \text{otherwise.} \end{cases}$$

This mimics the structure (2) very closely and, as with the random-effect bootstrap, for $j \neq l$, we can write $\operatorname{cov}^*(Y_{ij}^*, Y_{il}^*) = \operatorname{var}^*(Y_{ij}^*) - (mg)^{-1} S_{W2}$. However, for the two-stage bootstrap, the expected variance

$$E\{\mathrm{var}^*(Y_{ij}^*)\} = (mg)^{-1}\{m(g-1)\sigma_\beta^2 + (mg-1)\sigma_\varepsilon^2\}$$

and the expected covariance

$$\mathrm{cov}^*(Y_{ij}^*, Y_{il}^*) = g^{-1}(g-1)(\sigma_\beta^2 + m^{-1}\sigma_\varepsilon^2)$$

which are not correct unless both m and g tend to ∞. Thus we reach opposite conclusions from the bootstrap moments and the expected bootstrap moments. Extensive calculations show that the two-stage bootstrap variance of the within sum of squares and the two-stage bootstrap covariance of the between and within sums of squares are not consistent even when $m, g \to \infty$. Thus the fact that the observations have the correct low moment structure does not imply that the bootstrap will be consistent for a general statistic.

McCullagh (2000) also considered some further bootstraps which involve sampling without replacement or permutation. As we saw in discussing the cluster and randomized cluster bootstraps, permutation does not affect the between and within sums of squares because they are invariant to permutation. It does, however, affect the bootstrap moments and expected bootstrap moments of observations, undermining the value of criteria that are based on these moments.

Finally, it is worth noting that, if we specify the distribution of the random effect and residuals in model (1), we can construct parametric bootstraps. These bootstraps do produce consistent estimators of the variances and covariances that we require but only under the additional requirement that the distributions assumed are correct. This makes them of less interest in practice unless the distributional assumptions can be verified, and we do not consider them further.

4. Simulation

We carried out simulations for the random-effect, residual and cluster bootstrap for data that were generated by both the random-effect model and the transformation model. Although we have already shown in Table 1 that the theoretical target values for the variances and covariances of the sums of squares vary substantially both according to the model and to the generating distribution(s), it is useful to provide numerical results for small sample sizes. As we shall show, the numerical results demonstrate that the appropriate bootstrap gives reasonable results which are a substantial improvement over what would be obtained from normal theory when the normal is replaced by a t-distribution.

In each situation from Table 1, we generated 1000 data sets. For each data set, we computed the bootstrap variance for the total, S_{B2} and S_{W2} using 100 bootstrap replications. Table 2 gives the means and their standard deviation of these 1000 bootstrap variances for several of the cases. Results for the intermediate cases that are not shown typically lay between those given in Table 2. It is important to be cautious in interpreting the numerical results as the distributions of the bootstrap variance are very long tailed to the right, especially for the $t(5)$ distribution. For instance under the random-effect model with $t(5)$–$t(5)$ the distribution of the random-effect bootstrap variance has third quartile 1020 and maximum value 346600 with similar patterns for the other bootstraps under both models. The extreme skewness in the distributions suggests that, if we are going to construct confidence intervals, we should do so on the log-scale.

The random-effect bootstrap gives good results for the total and S_{B2} under the random-effect model. For the transformation model, it underestimates the S_{B2}-variance for the t-distribution. Across all models, the S_{W2} variance is underestimated for the t-distribution. The residual bootstrap is reliable for the total and for S_{B2} for the transformation model and the t-distribution. Its

Table 2. Results from simulations based on 1000 data sets with 100 bootstrap replicates per data set†

g = 5					
Random effect	N–N	Target	Random-effect bootstrap	Residual bootstrap	Cluster bootstrap
	Total	100	102 (2)	102 (2)	81 (2)
	Between	200	202 (8)	298 (14)	112 (5)
	Within	30	29 (1)	30 (1)	24 (1)
Transformation	N (1,1)	Target	Random-effect bootstrap	Residual bootstrap	Cluster bootstrap
	Total	100	98 (2)	100 (2)	79 (2)
	Between	200	204 (10)	293 (17)	113 (6)
	Within	30	30 (1)	31 (1)	25 (1)
Random effect	$t(5)$–$t(5)$	Target	Random-effect bootstrap	Residual bootstrap	Cluster bootstrap
	Total	167	177 (7)	183 (7)	141 (5)
	Between	1422	1284 (428)	1550 (359)	739 (236)
	Within	271	128 (7)	140 (10)	134 (10)
Transformation	$t^*(5)$ (5/3,5/3)	Target	Random-effect bootstrap	Residual bootstrap	Cluster bootstrap
	Total	167	163 (4)	170 (4)	130 (3)
	Between	889	620 (44)	1057 (91)	336 (23)
	Within	271	133 (13)	167 (16)	139 (17)
g = 15					
Random effect	$t(5)$–$t(5)$	Target	Random-effect bootstrap	Residual bootstrap	Cluster bootstrap
	Total	500	504 (11)	521 (11)	468 (9)
	Between	5483	4755 (1076)	3315 (280)	4148 (853)
	Within	813	442 (36)	463 (42)	552 (55)
Transformation	$t^*(5)$ (5/3,5/3)	Target	Random-effect bootstrap	Residual bootstrap	Cluster bootstrap
	Total	500	502 (7)	524 (8)	467 (7)
	Between	3306	2643 (153)	3576 (231)	2326 (156)
	Within	813	476 (45)	607 (64)	592 (68)

†Entries give the target values from Table 1 and the mean of the 1000 bootstrap estimates of the variance along with their standard deviation in parentheses. The distribution labels are as in Table 1.

performance for S_{B2} under normality is conservative but we note that we may be seeing difficulties due to the long tail. The medians of the bootstrap variance for S_{B2} are 149 and 124 for the two models. Again the bootstrap variance for S_{W2} is underestimated for the t-distribution but there is a slight improvement over the random-effect bootstrap. The cluster bootstrap consistently underestimates the bootstrap variance for all the statistics. However, we should note that for the cluster bootstrap we are essentially working with five and 15 observations respectively. In addition, one can argue that the cluster results should be scaled by $g/(g-1)$. However, the effect of this rescaling does not overcome the underestimation.

We observe that, for non-normal data, using any of the bootstraps gives a substantially better variance estimate than would be obtained under a normal approximation. For instance with S_{B2} and the random-effect model with $t(5)$–$t(5)$, the normal theory gives a variance estimate of about 200 with a target of 1422 for $g=5$. The three bootstraps give variance estimates of 1284, 1550 and 739, all moving well in the direction of the target.

5. Conclusions

We have considered two different models for describing clustered data and shown that, in the non-normal case, the between and within sum of squares have different properties under the two models. We also demonstrated varying behaviour for the bootstrap variances and covariances of the between and within sum of squares under different bootstrap schemes and the two models. Criteria based on the distribution of observations such as the moment criterion of Davison and Hinkley (1997) and the exchangeability criterion of McCullagh (2000) are not really useful for anticipating the properties of general statistics. Instead, we have focused directly on consistent variance estimation for a class of statistics. In this framework, the random-effect bootstrap gives consistent variance estimates under the random-effect model as the number of clusters and the size of each cluster increases. The residual bootstrap which is reminiscent of the usual bootstrap in regression models gives consistent variance estimates under the transformation model as the number of clusters increases. The cluster bootstrap provides a very simple resampling scheme which yields consistent variance estimates as the number of clusters increases for both the random-effect and the transformation model. However, the cluster bootstrap is not as easily generalized to more complicated random and mixed models as the random-effect and residual bootstraps.

Acknowledgements

This research has been supported by the Engineering and Physical Sciences Research Council (GR/S21977/01), Australian Research Council (DP0559135) and Natural Science and Engineering Research Council of Canada. We are grateful to the Joint Editor, Associate Editor and two referees for comments which improved this paper.

Apppendix A

A.1. Derivation of equation (5) for the covariance, cov($S_{\mathrm{B}2}$, $S_{\mathrm{W}2}$)

We have

$$
\begin{aligned}
E(S_{\mathrm{B}2}S_{\mathrm{W}2}) = {} & E\left[\left\{m\sum_{i=1}^{g}(\beta_i^2+2\beta_i\bar{\varepsilon}_{i\cdot}+\bar{\varepsilon}_{i\cdot}^2)-mg(\bar{\beta}^2+2\bar{\beta}\bar{\varepsilon}_{\cdot\cdot}+\bar{\varepsilon}_{\cdot\cdot}^2)\right\}\left(\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2-m\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\right)\right]\\
= {} & m\,E\left\{\sum_{i=1}^{g}(\beta_i^2+2\beta_i\bar{\varepsilon}_{i\cdot}+\bar{\varepsilon}_{i\cdot}^2)\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}-mg\,E\left\{(\bar{\beta}^2+2\bar{\beta}\bar{\varepsilon}_{\cdot\cdot}+\bar{\varepsilon}_{\cdot\cdot}^2)\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}\\
& -m^2\,E\left\{\sum_{i=1}^{g}(\beta_i^2+2\beta_i\bar{\varepsilon}_{i\cdot}+\bar{\varepsilon}_{i\cdot}^2)\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\right\}+m^2g\,E\left\{(\bar{\beta}^2+2\bar{\beta}\bar{\varepsilon}_{\cdot\cdot}+\bar{\varepsilon}_{\cdot\cdot}^2)\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\right\}\\
= {} & (mg)^2\sigma_\beta^2\sigma_\varepsilon^2+m\,E\left(\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right)-m^2g\sigma_\beta^2\sigma_\varepsilon^2-mg\,E\left(\bar{\varepsilon}_{\cdot\cdot}^2\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right)-mg^2\sigma_\beta^2\sigma_\varepsilon^2\\
& -m^2\,E\left\{\left(\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\right)^2\right\}+mg\sigma_\beta^2\sigma_\varepsilon^2+m^2g\,E\left(\bar{\varepsilon}_{\cdot\cdot}^2\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\right)\\
= {} & mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2+\frac{1}{m}E\left\{\sum_{i=1}^{g}\left(\sum_{j=1}^{m}\varepsilon_{ij}\right)^2\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}-\frac{1}{mg}E\left\{\left(\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}\right)^2\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}\\
& -m^2\frac{g-1}{g}E\left\{\left(\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^2\right)^2\right\}\\
= {} & mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2+\frac{1}{m}\,E\left\{\sum_{i=1}^{g}\left(\sum_{j=1}^{m}\varepsilon_{ij}\right)^2\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}-\frac{1}{mg}E\left\{\sum_{i=1}^{g}\left(\sum_{j=1}^{m}\varepsilon_{ij}^2+\sum_{j\neq k}\varepsilon_{ij}\varepsilon_{ik}\right)\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}\\
& -m^2\frac{g-1}{g}\,E\left(\sum_{i=1}^{g}\bar{\varepsilon}_{i\cdot}^4+\sum_{i\neq h}\bar{\varepsilon}_{i\cdot}^2\bar{\varepsilon}_{h\cdot}^2\right)\\
= {} & mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2+\frac{g-1}{mg}E\left\{\sum_{i=1}^{g}\left(\sum_{j=1}^{m}\varepsilon_{ij}\right)^2\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}-m^2\frac{g-1}{g}\left[g\frac{1}{m^4}\{m\,E(\varepsilon^4)\right.\\
& \left.+3m(m-1)\sigma_\varepsilon^4\}+\frac{g(g-1)}{m^2}\sigma_\varepsilon^4\right]\\
= {} & mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2+\frac{g-1}{mg}\,E\left\{\left(\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2+\sum_{i=1}^{g}\sum_{j\neq k}\varepsilon_{ij}\varepsilon_{ik}\right)\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\right\}-\frac{g-1}{m}E(\varepsilon^4)\\
& -(g-1)\left\{\frac{3(m-1)}{m}+g-1\right\}\sigma_\varepsilon^4
\end{aligned}
$$

$$
\begin{aligned}
&= mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2 + \frac{g-1}{mg}\, E\Big\{\Big(\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^2\Big)^2\Big\} - \frac{g-1}{m}E(\varepsilon^4) - \frac{g-1}{m}(2m-3+mg)\sigma_\varepsilon^4 \\
&= mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2 + \frac{g-1}{mg}\, E\Big(\sum_{i=1}^{g}\sum_{j=1}^{m}\varepsilon_{ij}^4 + \sum_{i=1}^{g}\sum_{j\neq k}^{m}\varepsilon_{ij}^2\varepsilon_{ik}^2 + \sum_{i\neq h}\sum_{j=1}^{m}\varepsilon_{ij}^2\varepsilon_{hj}^2 + \sum_{i\neq h}\sum_{j\neq k}^{m}\varepsilon_{ij}^2\varepsilon_{hk}^2\Big) \\
&\quad - \frac{g-1}{m}E(\varepsilon^4) - \frac{g-1}{m}(2m-3+mg)\sigma_\varepsilon^4 \\
&= mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2 + (g-1)E(\varepsilon^4) + (g-1)\{(m-1)+(g-1)+(m-1)(g-1)\}\sigma_\varepsilon^4 \\
&\quad - \frac{g-1}{m}E(\varepsilon^4) - \frac{g-1}{m}(2m-3+mg)\sigma_\varepsilon^4 \\
&= mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2 + \frac{(m-1)(g-1)}{m}E(\varepsilon^4) + (g-1)\Big\{(m-1)+(g-1)+(m-1)(g-1) \\
&\quad - \frac{1}{m}(2m-3+mg)\Big\}\sigma_\varepsilon^4 \\
&= mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2 + \frac{(m-1)(g-1)}{m}E(\varepsilon^4) + \frac{g-1}{m}(m-1)(mg-3)\sigma_\varepsilon^4
\end{aligned}
$$

so

$$
\begin{aligned}
\operatorname{cov}(S_{\mathrm{B2}}, S_{\mathrm{W2}}) &= E(S_{\mathrm{B2}}S_{\mathrm{W2}}) - E(S_{\mathrm{B2}})\,E(S_{\mathrm{W2}}) \\
&= mg(m-1)(g-1)\sigma_\beta^2\sigma_\varepsilon^2 + \frac{(m-1)(g-1)}{m}E(\varepsilon^4) + \frac{g-1}{m}(m-1)(mg-3)\sigma_\varepsilon^4 \\
&\quad - (m-1)g(g-1)(m\sigma_\beta^2\sigma_\varepsilon^2 + \sigma_\varepsilon^4) \\
&= \frac{(m-1)(g-1)}{m}\{E(\varepsilon^4) - 3\sigma_\varepsilon^4\},
\end{aligned}
$$

giving the result.

A.2. Derivation of the expression for the covariance cov(S^*_{B2}, S^*_{W2}) in theorem 5*

Note first that

$$
\begin{aligned}
0 &= \sum_{i=1}^{g} S_{\mathrm{W2}i}\Big\{\sum_{i=1}^{g}(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})\Big\}^2 \\
&= \sum_{i=1}^{g} S_{\mathrm{W2}i}(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})^2 + \sum_{i\neq h} S_{\mathrm{W2}i}(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot})^2 + 2\sum_{i\neq h} S_{\mathrm{W2}i}(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot}) + \sum_{i\neq h\neq k} S_{\mathrm{W2}i}(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot})(\bar Y_{k\cdot} - \bar Y_{\cdot\cdot}) \\
&= \frac{1}{m}S_{\mathrm{W2}}S_{\mathrm{B2}} - 2\sum_{i=1}^{g} S_{\mathrm{W2}i}(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})^2 + \sum_{i\neq h\neq k} S_{\mathrm{W2}i}(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot})(\bar Y_{k\cdot} - \bar Y_{\cdot\cdot})
\end{aligned}
$$

so

$$
\sum_{i\neq h\neq k} S_{\mathrm{W2}i}(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot})(\bar Y_{k\cdot} - \bar Y_{\cdot\cdot}) = 2\sum_{i=1}^{g} S_{\mathrm{W2}i}(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})^2 - \frac{1}{m}S_{\mathrm{W2}}S_{\mathrm{B2}}.
$$

Let Δ_i equal the number of times that cluster i appears in the sample, so $(\Delta_1, \ldots, \Delta_g) \sim \text{multinomial}(g, 1/g, \ldots, 1/g)$. Then,

$$
\begin{aligned}
E^*(S^*_{\mathrm{W2}}S^*_{\mathrm{B2}}) &= m\,E^*\Big\{\sum_{i=1}^{g}\Delta_i S_{\mathrm{W2}i}\sum_{i=1}^{g}\Delta_i(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})^2\Big\} - \frac{m}{g}\,E^*\Big[\sum_{i=1}^{g}\Delta_i S_{\mathrm{W2}i}\Big\{\sum_{i=1}^{g}\Delta_i(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})\Big\}^2\Big] \\
&= m\,E^*\Big\{\sum_{i=1}^{g}\Delta_i^2 S_{\mathrm{W2}i}(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})^2\Big\} + m\,E^*\Big\{\sum_{i\neq h}\Delta_i\Delta_h S_{\mathrm{W2}i}(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot})^2\Big\} \\
&\quad - \frac{m}{g}\,E^*\Big[\sum_{i=1}^{g}\Delta_i S_{\mathrm{W2}i}\Big\{\sum_{i=1}^{g}\Delta_i^2(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})^2 + \sum_{i\neq h}\Delta_i\Delta_h(\bar Y_{i\cdot} - \bar Y_{\cdot\cdot})(\bar Y_{h\cdot} - \bar Y_{\cdot\cdot})\Big\}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&= m\frac{2g-1}{g}\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 + m\frac{g-1}{g}\sum_{i\neq h} S_{W2i}(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})^2 - \frac{m}{g}E^*\Big\{\sum_{i=1}^{g}\Delta_i^3 S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2\Big\}\\
&\quad -\frac{m}{g}E^*\Big\{\sum_{i\neq h}\Delta_i\Delta_h^2 S_{W2i}(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})^2\Big\} - 2\frac{m}{g}E^*\Big\{\sum_{i\neq h}\Delta_i^2\Delta_h S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})\Big\}\\
&\quad -\frac{m}{g}E^*\Big\{\sum_{i\neq h\neq k}\Delta_i\Delta_h\Delta_g S_{W2i}(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})(\bar{Y}_{k\cdot}-\bar{Y}_{\cdot\cdot})\Big\}\\
&= m\frac{2g-1}{g}\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 + m\frac{g-1}{g}\sum_{i\neq h} S_{W2i}(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})^2 - \frac{m}{g^3}(5g^2-6g+2)\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2\\
&\quad -\frac{m}{g^3}(2g^2-4g+2)\sum_{i\neq h} S_{W2i}(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})^2 - 2\frac{m}{g^3}(2g^2-4g+2)\sum_{i\neq h} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})\\
&\quad -\frac{m}{g^3}(g^2-3g+2)\sum_{i\neq h\neq k} S_{W2i}(\bar{Y}_{h\cdot}-\bar{Y}_{\cdot\cdot})(\bar{Y}_{k\cdot}-\bar{Y}_{\cdot\cdot})\\
&= m\left(\frac{2g-1}{g}-\frac{g-1}{g}\right)\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 + \frac{g-1}{g}S_{W2}S_{B2} - \frac{m}{g^3}(5g^2-6g+2-2g^2+4g-2)\\
&\quad \times\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 - \frac{1}{g^3}(2g^2-4g+2)S_{W2}S_{B2} + 2\frac{m}{g^3}(2g^2-4g+2)\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2\\
&\quad -\frac{m}{g^3}(g^2-3g+2)\Big\{2\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 - \frac{1}{m}S_{W2}S_{B2}\Big\}\\
&= \frac{m(g-1)}{g}\sum_{i=1}^{g} S_{W2i}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 + \frac{(g-1)^2}{g^2}S_{W2}S_{B2}
\end{aligned}
$$

and the result follows.

References

Andersson, M. K. and Karlsson, S. (2001) Bootstrapping error component models. *Computnl Statist.*, **16**, 221–231.

Butar, F. B. and Lahiri, P. (2003) On measures of uncertainty of empirical Bayes small area estimators. *J. Statist. Planng Inf.*, **112**, 63–76.

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. *Appl. Statist.*, **52**, 431–443.

Davison, A. C and Hinkley, D. V. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Gross, S. (1980) Median estimation in sample surveys. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 181–184.

Lahiri, P. (2003) On the impact of bootstrap in survey sampling and small-area estimation. *Statist. Sci.*, **18**, 199–210.

McCarthy, P. J. and Snowden, C. B. (1985) The bootstrap and finite population sampling. *Vital and Health Statistics*, 2-95. Washington, DC: US Government Printing Office.

McCullagh, P. (2000) Re-sampling and exchangeable arrays. *Bernoulli*, **6**, 285–301.

Rao, J. N. K. and Wu, C.-F. J. (1988) Resampling inference with complex survey data. *J. Am. Statist. Ass.*, **83**, 231–241.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992) *Model Assisted Survey Sampling*. New York: Springer.

Scheffé, H. (1959) *The Analysis of Variance*. New York: Wiley.

Searle, S., Casella, G. and McCulloch, C. E. (1992) *Variance Components*. New York: Wiley.

Shao, J. (2003) Impact of the bootstrap on sample surveys. *Statist. Sci.*, **18**, 191–198.

Shao, J. and Tu, D. (1995) *The Jackknife and Bootstrap*. New York: Springer.

Sitter, R. (1992) A resampling procedure for complex survey data. *J. Am. Statist. Ass.*, **87**, 755–765.

Tukey, J. W. (1956) Variances of variance components: I, Balanced designs. *Ann. Math. Statist.*, **27**, 722–736.

Ukoumunne, O. C., Davison, A. C., Gulliford, M. C. and Chinn, S. (2003) Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. *Statist. Med.*, **22**, 3805–3821.

Welsh, A. H. (2005) Bootstrapping in clustered populations. *55th Sessn International Statistical Institute*, *Sydney*.