Statistics and Computing (1999) **9**, 3–8

Delete-m *Jackknife for Unequal* m

FRANK M. T. A. BUSING*, ERIK MEIJER† and RIEN VAN DER LEEDEN*

**Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, the Netherlands; e-mail: busing@rulfsw.fsw.leidenuniv.nl*

†Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, the Netherlands

Submitted October 1997 and accepted December 1998

In this paper, the delete-m_j jackknife estimator is proposed. This estimator is based on samples obtained from the original sample by successively removing mutually exclusive groups of unequal size. In a Monte Carlo simulation study, a hierarchical linear model was used to evaluate the role of nonnormal residuals and sample size on bias and efficiency of this estimator. It is shown that bias is reduced in exchange for a minor reduction in efficiency. The accompanying jackknife variance estimator even improves on both bias and efficiency, and, moreover, this estimator is mean-squared-error consistent, whereas the maximum likelihood equivalents are not.

Keywords: jackknife, Monte Carlo simulation, hierarchical linear model, (un)balanced samples, maximum likelihood estimation

1. Introduction

Consider a sample of size n from some distribution and an estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ of a parameter θ obtained from this sample. Suppose $\hat{\theta}_{n-m}$ is the same estimator with a group of m observations removed from the sample. The bias of $\hat{\theta}_n$ can be estimated from differences between $\hat{\theta}_n$ and various $\hat{\theta}_{n-m}$. Under suitable regularity conditions, the bias of the bias-corrected jackknife estimator $\hat{\theta}_J$ is of order n^{-2}, when m is relatively small compared to n, whereas the bias of $\hat{\theta}_n$ is of order n^{-1}.

Tukey (1958) conjectured that it would be natural to estimate the variance of $\hat{\theta}_n$ by the sample variance based on the pseudo values. Although it was originally an estimator of the variance of $\hat{\theta}_n$, and Efron (1982, p. 13) states that it is a better estimator of $\mathrm{var}(\hat{\theta}_n)$ than of $\mathrm{var}(\hat{\theta}_J)$, it is generally used as an estimator of the variance of $\hat{\theta}_J$.

There are many ways of partitioning a sample of size n into g mutually exclusive groups of size m. With n, g, and m all being integers, the number of groups g is equal to the total sample size n divided by the group size m, $g = n/m$. One special case exists where $m = 1$ and $g = n$. In practice, the sample may consist of existing groups, such as households or school classes. It seems logical to use this existing partitioning of the sample to form the g groups. However, the groups in the sample may have varying sizes. In that case, n/m_j (where m_j is the size of group j, $j = 1, \ldots, g$) may, for some j, not be equal to the number of groups g, and adaptions to current formulas are in order.

The jackknife was originally developed for i.i.d. data: the observations were assumed to be independent. Whether the jackknife was applied to single observations or groups of observations, the independence assumption was not violated as long as the single observations were independent. However, considering existing groups, dependence may be expected between observations within the same group. Nevertheless, when the groups are independent, the jackknife may be applied to the groups instead of the original observations (see, for example, Wolter, 1985, chapter 4; Shao & Tu, 1995, chapter 9).

In this paper, a jackknife estimator is proposed, which is based on subsamples obtained from the original sample by successively removing mutually exclusive groups of unequal size. Formulas for this estimator and the corresponding variance estimator, which generalize formulas by Wolter (1985, p. 184), are given in Section 2. In Section 3, the design of a Monte Carlo simulation study using a hierarchical linear model is described. The results of this simulation study are presented in Section 4. Conclusions are given in Section 5.

2. Jackknife

As stated in the introduction, there are many ways to divide the simple into g mutually exclusive groups of size m. The choice of m determines the specific expression for the jackknife estimator and the jackknife variance estimator.

2.1. *Delete-1 jackknife*

Suppose $\hat{\theta}_n$ is an estimator of θ based on a sample of size n. Let $\hat{\theta}_{(i)}$ be the estimator of θ with the i-th observation removed from the sample (based on a sample size of $n-1$). The delete-1 jackknife estimator of θ is given by

$$\hat{\theta}_{J(1)} = n\hat{\theta}_n - (n-1)\bar{\theta}_{(1)}, \tag{1}$$

where $\bar{\theta}_{(1)} = n^{-1}\sum_{i=1}^{n}\hat{\theta}_{(i)}$.

The jackknife variance estimator, based on the pseudo values

$$\tilde{\theta}_{(i)} = n\hat{\theta}_n - (n-1)\hat{\theta}_{(i)}, \quad i = 1, \ldots, n,$$

is given by

$$\begin{aligned}\hat{\sigma}^2_{J(1)} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-1}\left(\tilde{\theta}_{(i)} - \frac{1}{n}\sum_{k=1}^{n}\tilde{\theta}_{(k)}\right)^2, \\ &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \bar{\theta}_{(1)}\right)^2.\end{aligned} \tag{2}$$

2.2. *Delete-m jackknife*

Instead of removing one single observation from the sample, suppose the sample is divided into g mutually exclusive and independent groups of (equal) size m ($m > 1$), where $m = n/g$. Let $\hat{\theta}_{(j)}$ be the estimator of θ with the m observations of group j removed from the sample. The delete-m jackknife estimator of θ is given by

$$\hat{\theta}_{J(m)} = g\hat{\theta}_n - (g-1)\bar{\theta}_{(m)}, \tag{3}$$

with $\bar{\theta}_{(m)} = g^{-1}\sum_{j=1}^{g}\hat{\theta}_{(j)}$. For $m = 1$ this coincides with (1). Hence, $\hat{\theta}_{J(m)}$ is based on g estimators $\hat{\theta}_{(j)}$ of θ, each based on a subsample of size $n-m$.

The delete-m jackknife variance estimator, also analogous to the single deletion case (2), and based on the pseudo values

$$\tilde{\theta}_{(j)} = g\hat{\theta}_n - (g-1)\hat{\theta}_{(j)}, \quad j = 1, \ldots, g,$$

is given by

$$\begin{aligned}\hat{\sigma}^2_{J(m)} &= \frac{1}{g}\sum_{j=1}^{g}\frac{1}{g-1}\left(\tilde{\theta}_{(j)} - \frac{1}{g}\sum_{k=1}^{g}\tilde{\theta}_{(k)}\right)^2, \\ &= \frac{g-1}{g}\sum_{j=1}^{g}\left(\hat{\theta}_{(j)} - \bar{\theta}_{(m)}\right)^2.\end{aligned} \tag{4}$$

Derivations leading to (1) through (4) can be found in standard jackknife literature. Wolter (1985, section 4.6) states that formulas (3) and (4) can also be used in cluster sampling, when the data within clusters are dependent. A recent and systematic introduction to the theory of the jackknife, including a discussion of theoretical properties of the jackknife, can be found, for example, in Shao and Tu (1995).

2.3. *Delete-m_j jackknife*

Instead of using groups restricted to be of equal size, one could divide the sample into g groups of (possibly) different sizes. Let $\hat{\theta}_{(j^*)}$ be an estimator of θ based on a sample from which group j with size m_j is removed. The delete-m_j jackknife estimator of θ is given by

$$\hat{\theta}_{J(m_j)} = g\hat{\theta}_n - \sum_{j=1}^{g}\left(1 - \frac{m_j}{n}\right)\hat{\theta}_{(j^*)}.$$

This reduces to the delete-m jackknife estimator (3), when the group sizes are all equal.

The expression for $\hat{\theta}_{J(m_j)}$ can be justified as follows. Consider an estimator $\hat{\theta}$ of a parameter θ obtained from a sample of size n from some distribution. Many such estimators have an expected value that can be written as the true value plus a power series expansion in $1/n$, i.e.,

$$E\left[\hat{\theta}\right] = \theta + \frac{b_1}{n} + \frac{b_2}{n^2} + \frac{b_3}{n^3} + \cdots, \tag{5}$$

where θ is the true value and $b_1, b_2, \ldots$ are unknown constants, independent of sample size, and frequently not equal to zero (see, e.g., Quenouille, 1956; Schucany, Gray, & Owen, 1971). If $b_1 \neq 0$, the bias in (5) is clearly of order n^{-1}. For group j of size m_j, the total sample size can be written as $n = m_j h_j$.

Hence,

$$E\left[h_j\hat{\theta}\right] = h_j\theta + \frac{b_1}{m_j} + \frac{b_2}{h_j m_j^2} + \frac{b_3}{h_j^2 m_j^3} + \cdots, \tag{6}$$

and

$$\begin{aligned}E\left[\hat{\theta}_{(j^*)}\right] &= \theta + \frac{b_1}{(h_j-1)m_j} + \frac{b_2}{(h_j-1)^2 m_j^2} \\ &\quad + \frac{b_3}{(h_j-1)^3 m_j^3} + \cdots.\end{aligned} \tag{7}$$

Combining (6) and (7) gives

$$\begin{aligned}&E\left[h_j\hat{\theta} - (h_j-1)\hat{\theta}_{(j^*)}\right] \\ &= \theta + \frac{b_2}{m_j^2}\left(\frac{1}{h_j} - \frac{1}{h_j-1}\right) + \frac{b_3}{m_j^3}\left(\frac{1}{h_j^2} - \frac{1}{(h_j-1)^2}\right) + \cdots, \\ &= \theta - \frac{b_2}{n^2}\frac{h_j}{h_j-1} - \frac{b_3}{n^3}\frac{h_j(2h_j-1)}{(h_j-1)^2} + \cdots.\end{aligned}$$

Finally, to prevent loss of efficiency, the weighted average of the g possible estimators is used (cf. Quenouille, 1956), which gives

$$\hat{\theta}_{J(m_j)} = \sum_{j=1}^{g} \frac{1}{h_j}\left(h_j\hat{\theta}_n - (h_j-1)\hat{\theta}_{(j^*)}\right), \\ = g\hat{\theta}_n - \sum_{j=1}^{g}\left(1-\frac{m_j}{n}\right)\hat{\theta}_{(j^*)}. \tag{8}$$

The expectation of (8) is given by

$$E\left[\sum_{j=1}^{g}\frac{1}{h_j}\left(h_j\hat{\theta} - (h_j-1)\hat{\theta}_{(j^*)}\right)\right] \\ = \sum_{j=1}^{g}\frac{1}{h_j}\theta - \sum_{j=1}^{g}\frac{1}{h_j}\left(\frac{b_2}{n^2}\frac{h_j}{h_j-1}\right) \\ - \sum_{j=1}^{g}\frac{1}{h_j}\left(\frac{b_3}{n^3}\frac{h_j(2h_j-1)}{(h_j-1)^2}\right) + \cdots, \\ = \theta - \frac{b_2}{n^2}\sum_{j=1}^{g}\frac{1}{h_j-1} - \frac{b_3}{n^3}\sum_{j=1}^{g}\frac{2h_j-1}{(h_j-1)^2} + \cdots,$$

which is of order n^{-2}, if $b_2 \neq 0$, and when m_j is relatively small compared to n.

The estimator of the variance of $\hat{\theta}_{J(m_j)}$, based on the pseudo values

$$\tilde{\theta}_{(j^*)} = h_j\hat{\theta}_n - (h_j-1)\hat{\theta}_{(j^*)}, \quad jj^* = 1, \ldots, g,$$

is given by

$$\hat{\sigma}^2_{J(m_j)} = \frac{1}{g}\sum_{j=1}^{g}\frac{1}{h_j-1}\left(\tilde{\theta}_{(j^*)} - \hat{\theta}_{J(m_j)}\right)^2, \\ = \frac{1}{g}\sum_{j=1}^{g}\frac{1}{h_j-1}\left(h_j\hat{\theta}_n - (h_j-1)\hat{\theta}_{(j^*)} - g\hat{\theta}_n + \sum_{k=1}^{g}\left(1-\frac{m_k}{n}\right)\hat{\theta}_{(k^*)}\right)^2.$$

This reduces to the delete-*m* jackknife variance estimator (4), when the group sizes are all equal.

The formulas in this section can (also) be applied to cluster sampling, in which the data within groups may be dependent, but the groups are independent. This generalizes the formulas of Wolter (1985, section 4.6), who discusses the jackknife for cluster sampling with all group sizes equal.

3. Simulation

As the delete-m_j jackknife is most relevant for situations in which the sample consists of existing groups, and in which therefore cases may be dependent within groups, we will study the properties of the proposed estimators in such a situation by means of a Monte Carlo simulation study. An appropriate approach for analyzing data containing independent groups of unequal size is using hierarchical linear models, also called multilevel models (see, for example, Kreft & de Leeuw, 1998). Examples of typical (hierarchical) data suited for these models are data of family members nested within households (two levels), time points nested within students (repeated measures), or students nested within classes, with classes nested within schools (three levels). The simplest possible hierarchical model is the random effects ANOVA model, which is given by

$$y_{ij} = \gamma + u_j + \varepsilon_{ij}, \tag{9}$$

where for every level-1 unit i (typically individuals; $i = 1, \ldots, m_j$) within every level-2 unit j (typically groups; $j = 1, \ldots, g$), y_{ij} contains a value on a response variable, and u_j and ε_{ij} contain the level-2 residuals and level-1 residuals, respectively.

Concerning this model, Busing, Meijer, and Van der Leeden (1994) and Goldstein (1995) suggest that the jackknife can be used to obtain bias-reduced estimates and standard errors for all model parameters.

The parameters to be estimated in (9) are the fixed component γ (grand mean), and a set of variance components, the variance σ^2_u of u, and the variance σ^2_ε of ε.

For the simulation study, the true parameters were set to the following values: γ was set to 1 and σ^2_u and σ^2_ε were set to 2 and 8, respectively. To be more specific, the residuals (u and ε) were generated from a lognormal distribution with a skewness (μ_3) of 2, set in deviation of its mean (expectation equals 0) and unit normalized (variance equals 1). Multiplying the generated values by the square root of the true values of σ^2_u and σ^2_ε mentioned earlier, produced the applied residuals. The true values of the standard errors were estimated by the standard deviations of the parameter estimates. The number of groups g was set to 20 and the group sizes m_j were drawn from a uniform distribution with a minimum of 10 and a maximum of 30 ($m_j \sim$ uniform(10, 30)). In the simulation study, μ_3, g and m_j were varied in turn.

A modified version of the `MLA` program (Busing *et al.*, 1994) was used to provide full information maximum likelihood (FIML) and restricted maximum likelihood (REML) estimators of all model parameters as well as the delete-m_j jackknife estimators (JACK). The jackknife estimators (specifically $\hat{\theta}_{(j^*)}$, $j^* = 1, \ldots, g$) were based on the FIML estimators. The standard errors of the ML estimators were derived from standard maximum likelihood theory. The standard errors of the delete-m_j jackknife estimators are simply the square roots of the delete-m_j jackknife variance estimators. The `RANLUX` pseudo-random number generator (Lüscher, 1994) was used to obtain the random samples. The number of replications (r) in the Monte Carlo simulation study was set to 10000, which was expected to be sufficient for accurate results.

4. Results

The results of the simulation study are evaluated by studying the properties of the estimators. The extent to which an estimator possesses certain properties, determines whether one estimator is preferred to another estimator.

The first part of this paper elaborates on the property of unbiasedness. Relative bias ($\mathrm{RB} = r^{-1}\sum_{i=1}^{r}(\hat{\theta}_i - \theta)/\theta$) is used to assure a fair comparison between parameters of different magnitude. Obviously, in the ideal case, RB should be zero. Another way of comparing the estimators is by their (fixed or small sample) efficiency. The degree of efficiency can be expressed by the relative mean squared error of an estimator ($\mathrm{RMSE} = r^{-1}\sum_{i=1}^{r}[(\hat{\theta}_i - \theta)/\theta]^2$).

Figure 1 shows the boxplots of the relative bias for σ_u^2. The boxplot clusters are defined by the FIML estimator and the delete-m_j jackknife estimator and shown for different values of the skewness of the residuals (u and ε).

The delete-m_j jackknife estimator is less biased, and equally efficient, in comparison with the FIML estimator. Similar results are indicated by the trimmed means in the left part of Table 1. Both standard errors show more relative bias as the skewness of the residuals increases (Figure 2). Further, Table 1 shows that, although the delete-m_j jackknife standard error is less biased, the FIML standard error has smaller relative mean squared error.

Figures 3–4 compare the ML estimators to the delete-m_j jackknife estimator. The boxplots of the relative bias for σ_u^2 and σ_ε^2 are displayed for a different number of groups (g).

Besides the relative bias of the FIML estimator, which is larger for σ_u^2 ($g \leq 80$) (compare e.g. Searle, Casella, & McCulloch, 1992; Busing, 1993), the results for the delete-m_j jackknife estimator are similar to the ML estimators for both relative bias and relative mean squared error.

Additionally, the boxplots clearly illustrate that the estimators are all relative-mean-squared-error consistent, a large sample property, which states that the values of an estimator tend to get closer to the true value as the sample size increases, $\lim_{n\to\infty} E[\{(\hat{\theta}_n - \theta)/\theta\}^2] = 0$. The distributions of the estimators of σ_u^2 and σ_ε^2 depend on g and n, respectively (see e.g. Busing, 1993; Longford, 1993), which is clearly illustrated by Figure 3 and Figure 4.

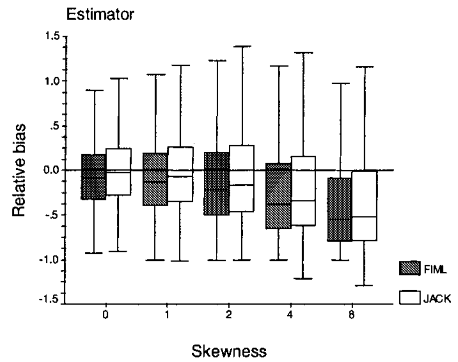

Fig. 1. *Boxplots of RB of σ_u^2 for different values of μ_3*

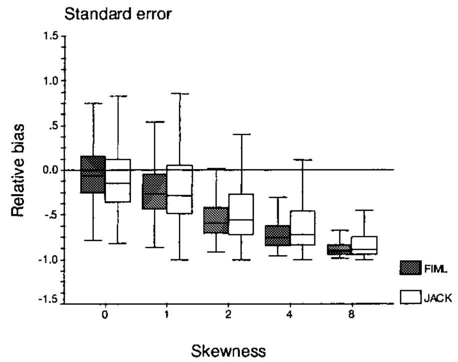

Fig. 2. *Boxplots of RB of SE(σ_u^2) for different values of μ_3*

The results for the corresponding standard errors are shown in Figures 5–6.

It is shown that the reported standard errors of the FIML estimators, derived from asymptotic theory and based on the assumed normal distribution of the estimator, may be quite different from the true standard errors. Although the variance of the ML standard errors become smaller as the sample size increases, the relative mean squared error does not converge to zero, because the

Table 1. *RB and RMSE of σ_u^2 and SE(σ_u^2) for different values of μ_3*

Skewness	Estimator				Standard error			
	RB		RMSE		RB		RMSE	
	FIML	JACK	FIML	JACK	FIML	JACK	FIML	JACK
0	−0.07	−0.01	0.09	0.10	0.22	0.15	0.15	0.16
1	−0.08	−0.03	0.13	0.13	0.04	0.09	0.10	0.24
2	−0.11	−0.05	0.23	0.24	−0.25	−0.10	0.16	0.33
4	−0.22	−0.17	0.34	0.35	−0.42	−0.20	0.28	0.49
8	−0.34	−0.29	0.48	0.50	−0.55	−0.29	0.40	0.63

Values represent 5% trimmed mean

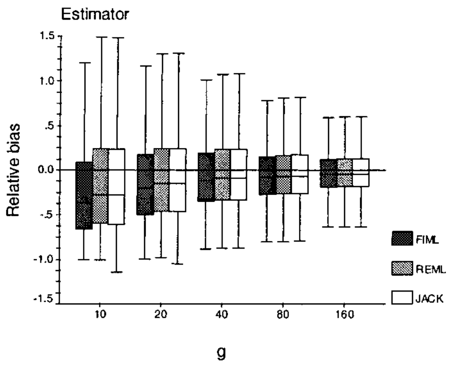

Fig. 3. *Boxplots of RB for σ_u^2 for different values of g*

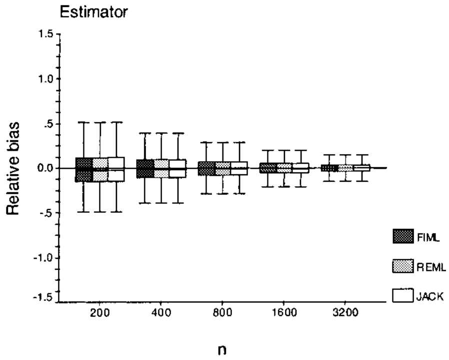

Fig. 4. *Boxplots of RB of σ_ε^2 for different values of $n \approx gm_j$*

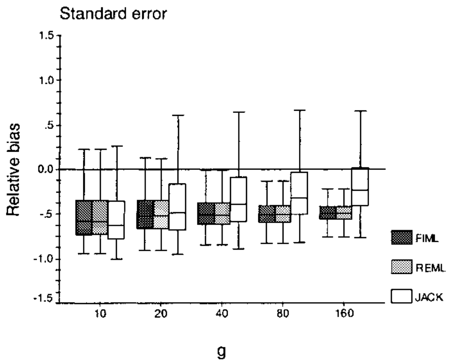

Fig. 5. *Boxplots of RB of $SE(\sigma_u^2)$ for different values of g*

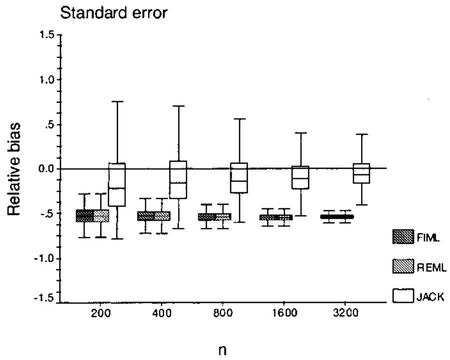

Fig. 6. *Boxplots of RB of $SE(\sigma_\varepsilon^2)$ for different values of $n \approx gm_j$*

standard errors remain biased. The ML standard errors are therefore not relative-mean-squared-error consistent. The delete-m_j jackknife standard errors are generally less biased, and as the sample size increases, the efficiency also increases, indicating relative-mean-squared-error consistency.

The simulation study also investigated the role of balancedness. Group sizes m_j were drawn from uniform distributions with increasing differences between the minimum and maximum of the distributions. The influence of (un)balancedness was however negligible, and where differences did occur, the results of both ML estimators and standard errors were comparable to the delete-m_j jackknife estimators and standard errors.

The fixed component (γ) was also evaluated and not displayed in the results. Magnus (1978) proved that maximum likelihood estimators of these components are unbiased. Reporting the results was redundant, because the relative bias and the relative mean squared error were nearly zero for both ML estimators and standard errors as well as for the delete-m_j jackknife estimator and standard error.

5. Conclusions

In this paper, the delete-m_j jackknife was introduced, and evaluated for the estimation of a hierarchical linear model. The delete-m_j jackknife applies the jackknife theory in a natural and theoretically correct manner on data consisting of independent groups of unequal size.

The delete-m_j jackknife estimators show a minor reduction in bias, in exchange for a comparable decrease in efficiency. The delete-m_j jackknife standard errors offer a distinct reduction in bias, and, in case of an adequate sample size, even an increase in efficiency. In any case, the delete-m_j jackknife standard errors are consistent, whereas the ML standard errors are not. Therefore, the jackknife may be preferred for inferential purposes.

Acknowledgements

The authors wish to thank Willem J. Heiser for his valuable comments on the earliest version of the paper. We would also like to express our thanks to the editor and the anonymous referees for their helpful suggestions and comments which led to improvements of the paper (and yet another 1.480.000.000 multilevel analyses).

References

Busing, F. M. T. A. (1993) *Distribution characteristics of variance estimates in two-level models; A Monte Carlo study* (Tech. Rep. No. PRM 93-04). Leiden, The Netherlands: Leiden University, Department of Psychology.

Busing, F. M. T. A., Meijer, E. and Van der Leeden, R. (1994) *MLA*, software for multilevel analysis of data with two levels. User's guide for version 1.0b (Tech. Rep. No. PRM 94-01). Leiden, The Netherlands: Leiden University, Department of Psychology.

Efron, B. (1982) *The jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.

Goldstein, H. (1995) *Multilevel statistical models* (2nd ed.). London: Edward Arnold.

Kreft, I. and de Leeuw, J. (1998) *Introduction to multilevel modelling*. London: Sage.

Longford, N. T. (1993) *Random coefficient models*. Oxford, GB: Clarendon Press.

Lüscher, M. (1994) A portable high-quality random number generator for lattice field theory simulations. *Computer Physics Communications*, **79**, 100–110.

Magnus, J. R. (1978) Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics*, **7**, 281–312.

Quenouille, M. H. (1956) Notes on bias in estimation. *Biometrika*, **43**, 353–360.

Schucany, W. R., Gray, H. L. and Owen, D. B. (1971) On bias reduction in estimation. *Journal of the American Statistical Association*, **66** (335), 524–533.

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

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

Tukey, J. W. (1958) Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, **29**, 614.

Wolter, K. M. (1985) *Introduction to variance estimation*. New York: Springer.