

**Multilevel Models Project**

**Working Paper  
December, 98**

**Consistent estimators for multilevel generalised  
linear models using an iterated bootstrap**

by

Harvey Goldstein  
hgoldstn@ioe.ac.uk

## Introduction

Several papers have addressed the issue of the parameter biases which can occur when fitting multilevel models with non Normal responses. Breslow and Clayton (1993) discuss various fitting procedures including those based upon linearising transformations, maximum likelihood and Bayesian estimation using MCMC. Direct maximum likelihood or restricted maximum likelihood, while feasible for simple models, becomes quickly intractable as the number of random effects increases: MCMC via Gibbs sampling is an attractive alternative, but the choice of prior distribution for the random parameters is important and there are difficulties in choosing 'diffuse' or uninformative priors (Browne, 1998?). Approximate methods based upon linearising transformations and applying quasiliikelihood estimation are attractive since they pose no serious computational problems and can be fitted using modifications to existing multilevel software packages.

Rodriguez and Goldman (1995) illustrate how severe underestimation can occur in a simple variance components model with binary responses, especially for the level 2 variance. They use a 'first order MQL' method (Goldstein, 1991). Goldstein (1995) and Goldstein and Rasbash (1996) develop improved linearising approximations and show that for models where there are adequate numbers of level 1 units per level 2 unit these give satisfactory results. Nevertheless, where the numbers of level 1 units per level 2 unit is small and for binary responses as in the Rodriguez-Goldman data sets, there is still some underestimation. In this paper we set out a procedure (Kuk, 1995) which yields asymptotically unbiased and consistent estimates for such models and which can be applied in general to any kind of non-linear multilevel model.

## Iterative bootstrap (IB) bias correction

We shall illustrate the procedure with a simple 2-level variance components model, as follows

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 x_{ij} + u_j$$

$$u_j \sim N(0, \sigma_u^2)$$

$$y_{ij} \sim \text{Binomial}(1, \pi_{ij})$$

Given a set of initial estimates, obtained using for example the first order MQL approximation,

$$\hat{\sigma}_u^{2(0)}, \hat{\beta}_0^{(0)}, \hat{\beta}_1^{(0)} \tag{1}$$

we generate a set of bootstrap samples, from the model using the estimates (1) and averaging over these we obtain the set of bootstrap estimates

$$\tilde{\sigma}_u^{2(0)}, \tilde{\beta}_0^{(0)}, \tilde{\beta}_1^{(0)} \tag{2}$$

We now obtain the bootstrap estimate of the bias by subtracting (2) from (1). These bias estimates are added to the initial parameter estimates (1) as a first adjustment to give new bias-corrected estimates

$$\hat{\sigma}_u^{2(1)}, \hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)} \quad (3)$$

We generate a new set of bootstrap samples from the model based upon the estimates given by (3), subtract the new mean bootstrap parameter estimates from (3) to obtain updated bias estimates and add these to the initial estimates (1) to obtain a new set of bias corrected estimates. When it converges, Kuk (1995) demonstrates that this procedure gives asymptotically consistent and unbiased parameter estimates.

In the present case the bootstrap samples have been generated parametrically by sampling from the distributions with estimated parameters: in the present case from a Normal distribution for the level 2 residuals and a binomial distribution (with denominator one) for the level 1 residuals. It relies upon the assumed model structure correctly representing the data hierarchy. In some cases this may not be the case, for example if an important level is omitted. Thus, the procedure does not protect against such forms of model misspecification. An important case is with discrete response models where we may have, say, extra binomial variation. In such cases the procedure can give different solutions depending on which estimation method is used.

Care needs to be taken with small variance estimates. To estimate the bias we need to allow negative estimates of variances. If an initial estimate is zero, then clearly, resetting negative bootstrap sample means to zero implies that the bias estimate will never be negative, so the new updated estimate will remain at zero. Moreover, as confirmed by simulations, all the estimates will exhibit a downward bias if negative bootstrap means are reset to zero. We also note that where an unbiased variance estimate is close to zero, the value of the bias is anyway small, so that full bias correction is less important and, for example, a second order PQL estimate may be adequate (see below).

The bootstrap replicates from the final bootstrap set generally will have too small a variance and so cannot directly be used for inference. If we knew the functional relationship between the bias-corrected value and the biased value this could be used to transform each of the bootstrap replicate estimates and the transformed values then used for inference. We shall discuss a procedure for doing this below. In *MLwiN* version 1.0 the procedure is to use scaling factors for each parameter calculated as follows.

For each parameter in turn, using the final bias-corrected estimate and the final bootstrap replicate mean, we take the ratio of these and multiply all the final replicate parameter values by this ratio. These scaled values are used to construct approximately correct standard errors and quantile estimates.

## A simulation

We simulate 100 replications of the model (1) for a binary (0,1) response with all three parameters equal to 1., with 50 level 2 units and 2 level 1 units per level 2 unit. This is a rather extreme case where we would expect serious underestimation of parameters.

To decide how many bootstrap samples we need for each iteration of the procedure we keep a running mean such that when, at the  $t$ -th bootstrap sample, for the running means  $\theta_t, \theta_{t-1}, \theta_{t-2}$

$$|\theta_t - \theta_{t-1}| < \varepsilon \quad \text{and} \quad |\theta_{t-1} - \theta_{t-2}| < \varepsilon \quad (4)$$

then we accept convergence. We have chosen the value of  $\varepsilon$  as 0.001 and set a minimum number of samples as 10. We note, in passing, that the device of maintaining a suitable running statistic to judge convergence is applicable for bootstrap sampling when attention is focused on other functions of parameters, for example the standard deviation or a percentile estimate. We then need a criterion for judging convergence of the bootstrap bias corrected estimates. In an application convergence needs to be monitored closely, especially for small values of random parameters. We finally adopted the following criteria for the simulations

We compute the average of the current and previous two estimates, say  $\bar{\theta}_t$ , and the average of the three estimates prior to these, say  $\bar{\theta}_{t-1}$ , and judge convergence as follows

$$\begin{aligned} |(\bar{\theta}_t - \bar{\theta}_{t-1}) / \bar{\theta}_t| &< 0.02 \quad \text{if} \quad \bar{\theta}_t \geq 0.25 \\ |(\bar{\theta}_t - \bar{\theta}_{t-1})| &< 0.005 \quad \text{if} \quad \bar{\theta}_t < 0.25 \end{aligned} \quad (5)$$

For small estimated values convergence is often slow and an absolute rather than relative criterion seems appropriate. The mean number of iterations required was 13.8 and the mean number of bootstrap samples per iteration was 80.5.

The basic results are given in Table 1. We have used the standard deviation rather than the variance for reporting means since the distribution of the latter is more skew.

<b>Table 1. Simulation results for MQL , Iterated bootstrap (IB)+ PQL estimates (s.e.)</b>						
	Level 2 s.d.		Intercept		Slope	
	Initial	IB	Initial	IB	Initial	IB
1st order MQL (IGLS)	0.49 (0.03)	0.98 (0.06)	0.89 (0.03)	1.05 (0.04)	0.91 (0.03)	1.07 (0.04)
1st order PQL (RIGLS)	0.49 (0.04)		0.88 (0.03)		0.88 (0.03)	
2nd order PQL (IGLS)	0.84 (0.06)		1.03 (0.04)		1.02 (0.03)	
2nd order PQL (RIGLS)	0.93 (0.07)		1.07 (0.04)		1.10 (0.04)	
The standard errors are computed over simulation replications.						

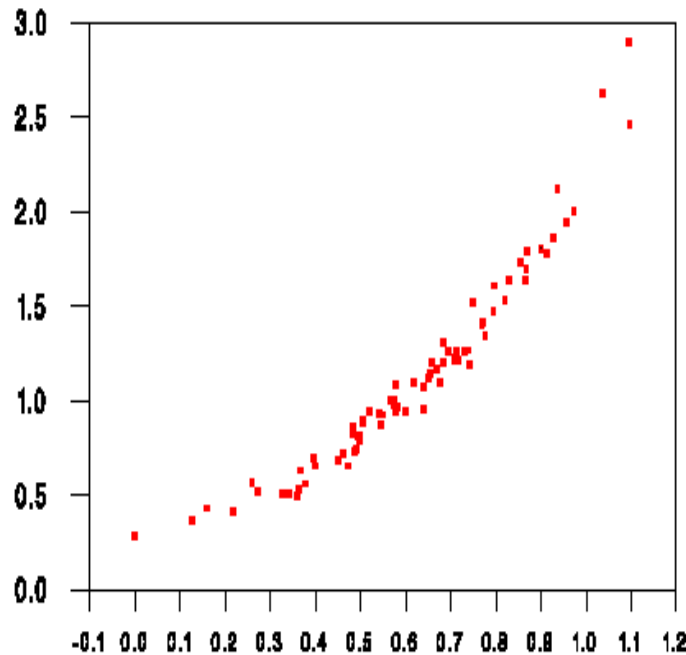
It is clear that the serious underestimation for all the parameters has been eliminated, and the final estimates are unbiased within the limits of sampling error. The initial second order PQL estimates using Iterative Generalised Least Squares (IGLS, which is maximum likelihood in the multivariate Normal case) of the fixed parameters in fact show no bias, but there is underestimation of the standard deviation. With Restricted Iterative Generalised Least Squares (RIGLS) which is restricted maximum likelihood in the multivariate Normal case) the variance estimate is less biased, although there appears to be a slight overestimation of the slope parameter. Interestingly, the first order PQL (RIGLS) estimates are no better than the first order MQL (IGLS) estimates, which suggests that second order PQL estimates should be used where possible for exploratory purposes. We also notice that the ratios of standard errors for the IB and MQL 1 estimates is approximately the same as the ratios of the parameter estimates, lending support to the scaling procedure suggested above.

It would of course be possible to start with the second order PQL estimates and use this estimation procedure for the bootstrapping. A difficulty with this is that each estimation takes rather longer and this will usually be an important consideration. Secondly, in some cases (5% in the present case) the second order procedure fails to converge whereas the first order one almost always does. We note, however, that discarding those replicates where convergence fails does not invalidate the IB procedure.

At convergence we generate a final sequence of bootstrap samples to provide estimates of precision, confidence intervals etc. The number of samples required for such purposes will generally be larger than used to in the updating, but as pointed out above we can use a running statistic for judging convergence at any prespecified accuracy.

Figure 1 shows the relationship between the final and initial estimates and illustrates how substantial adjustments can be made when the initial estimates are moderately large.

**Final estimate**



**Initial estimate**

**Figure 1. Final iterative bootstrap estimate of level 2 standard deviation by initial estimate. The value for the initial estimate of zero is the mean over the 22 such values.**

### **Interval estimation**

Once convergence has been achieved a final group of replicates can be produced as the basis for inference. As pointed out above, however, these generally will have too small variation. One solution would be to take every replicate set and use the IB to produce bias-corrected estimates; these could then directly be used for inference. This procedure, however, is too computationally intensive to be practical in most circumstances. Note that we cannot just bias correct for selected percentiles since the rank orders will differ among the parameters.

An alternative procedure is as follows, but it applies just to the random parameters. For each replicate in the final group we will have simulated a set of residuals from the assumed underlying multivariate normal distribution. Using the generated residuals we can obtain the empirical covariance matrix at each level of the model. Each element of this matrix (termed a generated parameter) corresponds to a random parameter estimate for the replicate and

we use the relationship between these two sets for our functional transformation. We note that this also allows us to establish functional relationships for any *function* of the random parameters. A suitable smoothing curve, such as a cubic spline, for relating the generated parameters to the estimated parameters is then required. By making the replicate set large enough we can obtain any required accuracy. This procedure does not deal with the fixed parameters. Here, however, the simple scaling procedure may be adequate, and the PQL2 estimates are typically almost unbiased.

This procedure can also be used to speed up the iterations - an accelerated iterated bootstrap. Consider the first replicate set. For a given parameter, if the distribution of the estimates covers the initial sample estimate ( $\hat{\theta}^{(0)}$ ) then the relationship between the generated parameter as response and the estimate obtained at that replicate allows us to obtain a predicted unbiased estimate. If this is not the case at the first iteration then we continue until it occurs. Using this estimate of the parameter we then iterate for a few further replicate sets to obtain an accurate unbiased estimate. From the final replicate set we then obtain the relationship to be used for inference.

## Conclusions

The procedure outlined is quite general, and can be applied to any non-linear multilevel model. As mentioned above, it will usually not be necessary where there are sufficient level 1 units per level 2 unit. In practice, where the number of such units is small, a useful strategy is to base model exploration on the second order (RIGLS) PQL estimates and then compute final bias corrected estimates using the first order MQL as here. In many cases, however, the second order (RIGLS) PQL estimates will be perfectly adequate.

Criteria are required for judging convergence and the number of bootstrap samples and the optimum criteria will generally depend on the data themselves and further work on this would be useful. For the bias corrected estimates the procedure may not always converge or convergence may be extremely slow. For MQL estimation neither of these problems has been encountered but they seem more likely to occur with PQL estimation and is a further reason for preferring the former to the latter.

## References

- Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalised linear models. *J. American Statistical Association*, 88, 9-25
- Goldstein, H. (1991). Non-linear multilevel models with an application to discrete response data. *Biometrika*, 73, 43-51.
- Goldstein, H. (1995) *Multilevel Statistical Models*. London, Edward Arnold; New York, Halstead Press.
- Goldstein, H. and Rasbash, J. (1996).. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, A*, **159**: 505-13
- Kuk, A.Y.C. (1995) Asymptotically unbiased estimation in generalised linear models with random effects. *J. Royal Statistical Society, B*, 57, 395-407
- Rodriguez, G. and Goldman, N (1995). An assessment of estimation procedures for multilevel models with binary responses. *J. Royal Statistical Society, A*, 158, 73-90