Weighting in Multilevel Models

Bing Tong

Introduction

Large-scale data sets usually use complex sampling design such as unequal probabilities of selection, clustering to collect data to save time and money. This leads to the necessity to incorporate sampling weights into multilevel models in order to obtain accurate estimates and valid inferences. However, the weighted multilevel models have been lately developed and minimal guidance is left on how to use sampling weights in multilevel models and which method is most appropriate.

In this paper, the following research questions are addressed:

1. How do MPML estimators differ from unweighted estimator in multilevel models in the informative and non-informative sampling designs in terms of relative bias, root mean square error and 95% confidence interval coverage rate?
2. How does intraclass correlation influence the performance of estimators under the informative and non-informative condition in terms of relative bias, root mean square error and 95% confidence interval coverage rate?

Theoretical Background

Multilevel Pseudo-Maximum Likelihood (MPML) Estimation Method

MPML methods with different scaling techniques are used in two-level random intercept models.

The MPML estimates $\hat{\theta}$ = $(\hat{\theta}_1, \hat{\theta}_2)$ are defined as the parameters to be estimated for the fixed effects for two levels respectively, and the population likelihood function is directly estimated by weighting the sampling likelihood function,

$$L(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \prod_{j=1}^{m} \left(\int (\prod_{i=1}^{n_j} f\left(y_{ij} \middle| x_{ij}, u_j, \boldsymbol{\theta_1}\right)^{w_{i|j}\lambda_{2j}}\right) \phi(u_j|z_j, \boldsymbol{\theta_2}) du_j)^{w_j\lambda_{1j}}$$

where u_j is the cluster-specific random effect; x_{ij} is individual-level covariates and z_j the cluster level covariates; $f\left(y_{ij} \middle| x_{ij}, u_j, \boldsymbol{\theta_1}\right)$ is the density function of the response variable y_{ij} and $\phi(u_j|z_j, \boldsymbol{\theta_2})$ the density function of u_j; w_j and $w_{i|j}$ are the cluster-level and individual-level weights respectively, λ_{1j} and λ_{2j} the scaling factors for the cluster-level and individual level sampling weights, respectively.

Scaling

Scaling is the primary tool for bias reduction. Two scaling techniques are introduced in the present study: effective cluster scaling and cluster scaling method. Pfeffermann et al. (1998) specified them as follows

$$\lambda_j = \frac{\sum_{i=1}^{n_j} w_{i|j}}{\sum_{i=1}^{n_j} w_{i|j}^2},$$

$$\lambda_j = \frac{n_j}{\sum_{i=1}^{n_j} w_{i|j}} \quad .$$

where n_j is the number of sample units in *j*th cluster.

Methods

Monte Carlo simulations are applied and the simulation design is as follows:

Table 1. *Simulation Design*

design	ICC	UW	RW	CS	ES
	ICC=0.5				
	ICC=0.3				
Informative	ICC=0.2				
	ICC=0.1				
	ICC=0.01				
	ICC=0.5				
	ICC=0.3				
Non-Informative	ICC=0.2				
	ICC=0.1				
	ICC=0.01				

The data are generated using the following model:
y_{ij} = 17.43 + 0.91*Female + 1.06*SES + 0.92*Pretest + 1.04*Rural + u_j + ε_{ij}.

Following Cai (2013), Asparouhov (2006) and Koziol et al. (2017), Poisson sampling is used to select schools and individuals. The *j*th school is selected with probability:

$$prob\,(I_j = 1) = \frac{1}{1+\exp\,(-\frac{\tilde{u}_{0j}}{2}+4.02)} \; .$$

The *i*th student within the selected *j*th school with probability:

$$prob\,(I_{i|j} = 1) = \frac{1}{1+\exp\,(-\frac{\tilde{e}_{ij}}{2}+1.23)}$$

where the $\tilde{u}_{0j}$ and $\tilde{e}_{0j}$ equal u_{0j} and e_{0j} respectively, and are rescaled to have a variance of 2.

Under the non-informative sampling condition, *j*th school is selected with probability of

$$prob\,(I_j = 1) = \frac{1}{1+\exp\,(-\frac{\beta_{0j}}{2}+4.02)} \; .$$

The *i*th student in the *j*th school is selected with probability of

$$prob\,(I_{i|j} = 1) = \frac{1}{1+\exp\,(-\frac{r_{ij}}{2}+1.23)}$$

where β_{0j} ~ N (0, 2) and r_{ij} ~ N (0, 2) and are not related to any variables in the model.

The simulation procedure is repeated 1000 times. Relative bias, root mean square error, and 95% confidence interval coverage rate are used to evaluate the quality of the estimators performance.

Results

Results for intercept and variance components estimation are reported here.

Substantial differences are found among these four estimation methods. In terms of bias, in the informative design, Figure 1 shows that the weighted estimators performs equally well and better than the unweighted for the intercept, whereas the cluster scaling estimator performs the best for the student-level variance. The unweighted estimator works the best for school-level variance estimation. In the non-informative design, Figure 2 indicates that the unweighted estimator performs the best or the second best for all of them.

In terms of RMSE, Figure 3 shows that including sampling weights decreases it for the intercept and student-level variance and increases it for the school-level variance in the informative design. However, Figure 4 shows that it increases the RMSE for all of them in the non-informative design. Therefore, the unweighted method works the most efficiently in the non-informative design.

Figure 5 shows that the weighted methods give better coverage rates for the intercept and student-level variance, but unweighted method does for school-level variance in the informative design. In the non-informative setting, Figure 6 shows that the unweighted method gives the best coverage rate for all the parameter estimates.

Tentatively, the cluster scaling estimator and effective scaling estimator might be preferred in the informative condition, while unweighted estimator does in the non-informative design.

ICC is one of the factors influencing the estimation quality (e.g., Asparouhov & Muthén, 2006; Kovačević & Rai, 2003). Figure 1-6 show that, the ICC affects relative bias and RMSE, but is not sensitive to coverage rate. As the ICC increases, the bias for student-level variance increases and the bias for school-level variance decreases in both conditions. These changes are obvious for school-level variance, but not for student-level variance. No monotonic patterns for the relative bias are found for the intercept in the informative condition, but clear patterns are in the non-informative design.

RMSE shows the similar patterns in both conditions for all the parameters except the unweighted estimator for the intercept. As the ICC increases, the RMSE decreases for the student-level variance, and increases for the school-level variance with all the four estimators.

Discussions and Future Research

Not all previous findings are confirmed in the current study. It is inferred that checking the informativeness of sampling design is necessary, because it will determine whether sampling weights should be employed. Second, researchers should examine the ICC and evaluate the magnitude and significance of variance components to determine whether multilevel modeling is necessary.

There are several limitations in this study. The primary limitation is that only a simple linear random-intercept model is applied. Second, different techniques can be adopted for weights, for example, trimming. Third, more levels of informativeness can be considered to offer a clearer picture under which condition, the estimates are biased. Above all, future research is needed to enhance weighted multilevel models.

References:

Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics—Theory and Methods*, *35*(3), 439-460.

Asparouhov, T., & Muthén, B. (2006). *Multilevel modeling of complex survey data.* Paper presented at the Proceedings of the Joint Statistical Meeting in Seattle.

Cai, T. (2013). Investigation of ways to handle sampling weights for multilevel model analyses. *Sociological Methodology*, *43*(1), 178-219.

Kovačević, M. S., & Rai, S. N. (2003). A pseudo maximum likelihood approach to multi-level modeling of survey data. *Communications in Statistics-Theory and Methods*, *32*(1), 103-121.

Koziol, N. A., Bovaird, J. A., & Suarez, S. (2017). A comparison of population-averaged and cluster-specific approaches in the context of unequal probabilities of selection. *Multivariate Behavioral Research*, *52*(3), 325-349. doi: 10.1080/00273171.2-17.1292115

Pfeffermann, D., Skinner, C. J., Holmes D. J., Goldstein, H. & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of Royal Statistical Society*: *Series B*, *60*(1), 23-40.

Appendices: Figures

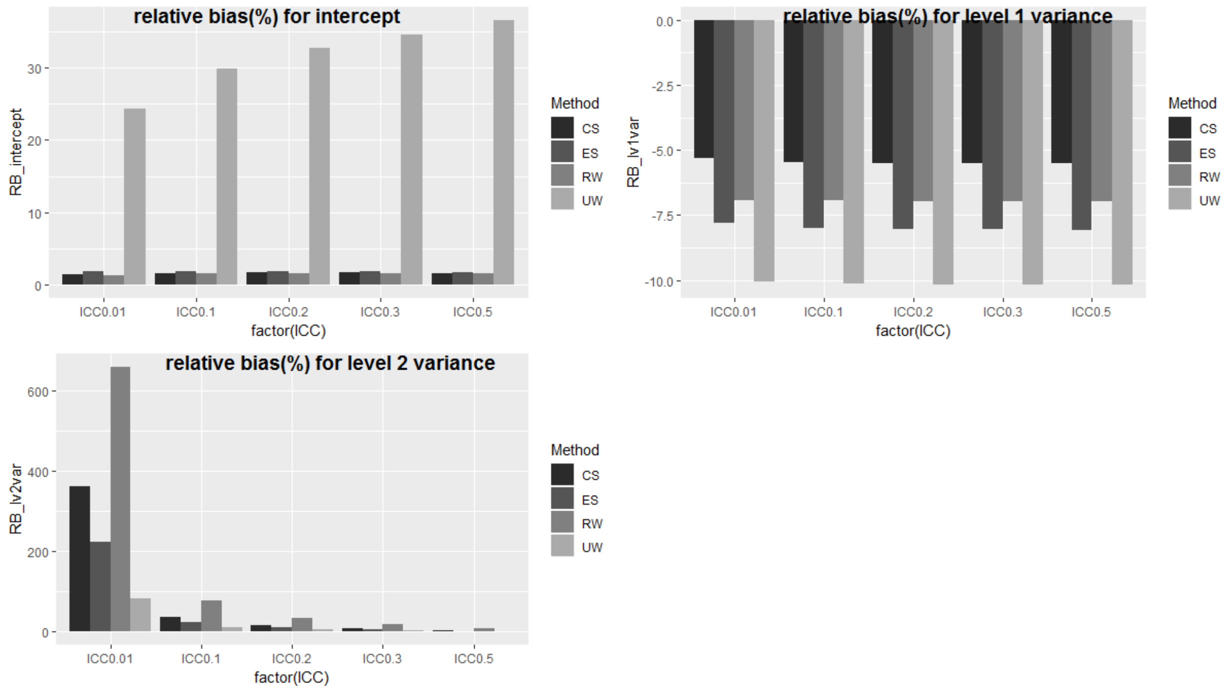

Figure 1. Relative bias (%) for intercept and variance components in the informative design

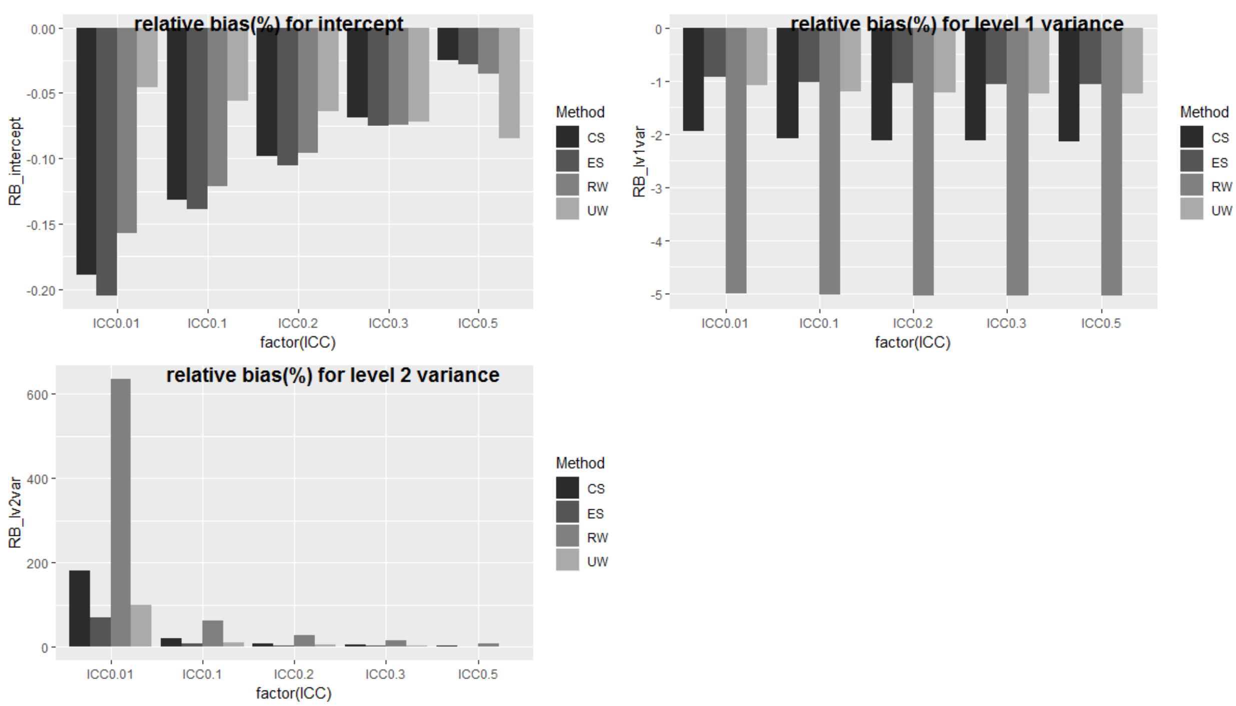

Figure 2. Relative bias (%) for intercept and variance components in the non-informative design

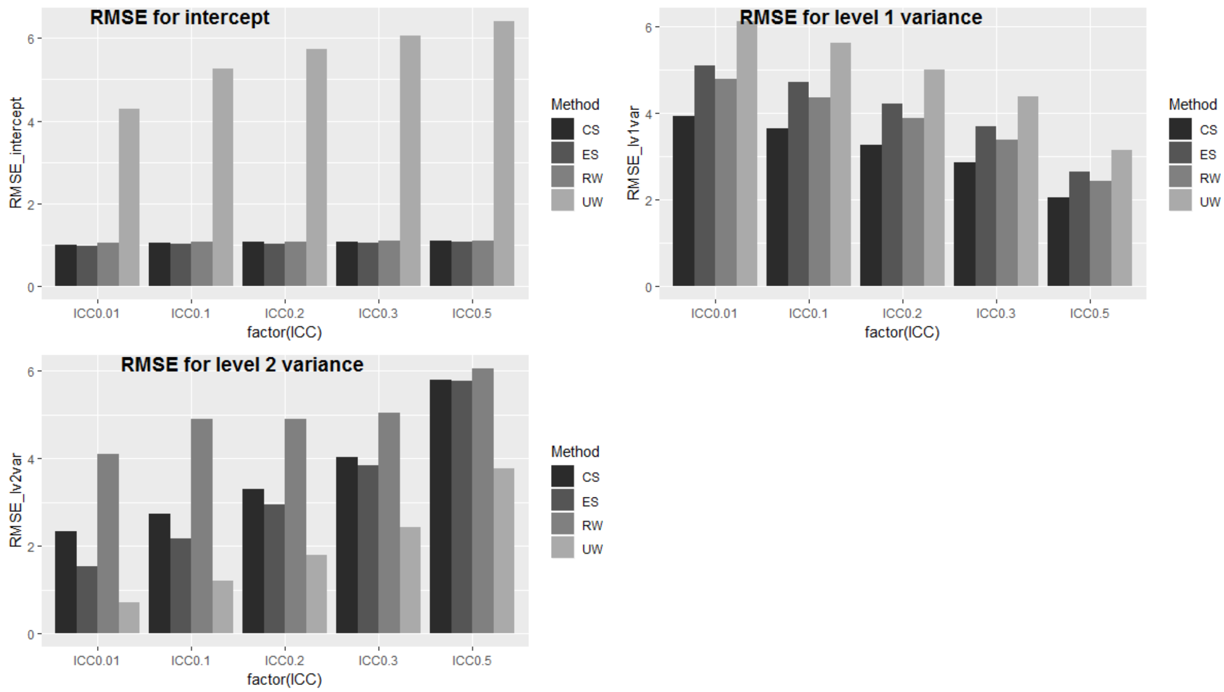

Figure 3. RMSE for intercept and variance components in the informative design

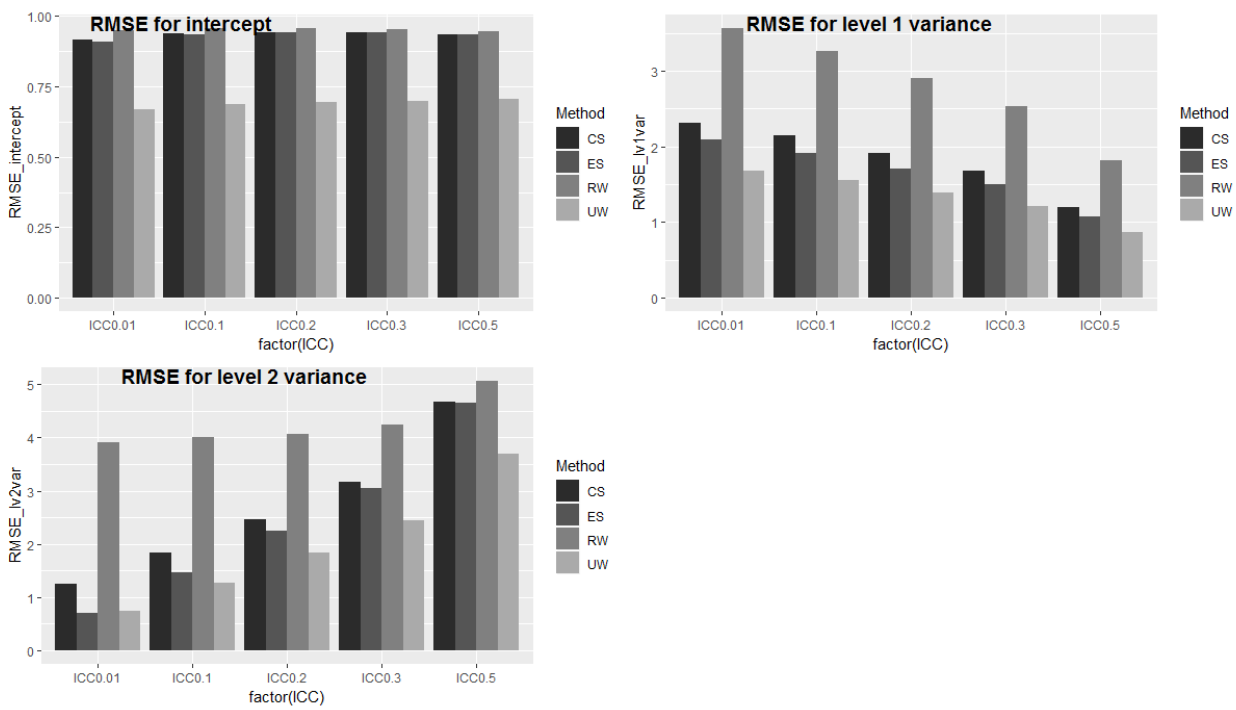

Figure 4. RMSE for intercept and variance components in the non-informative design

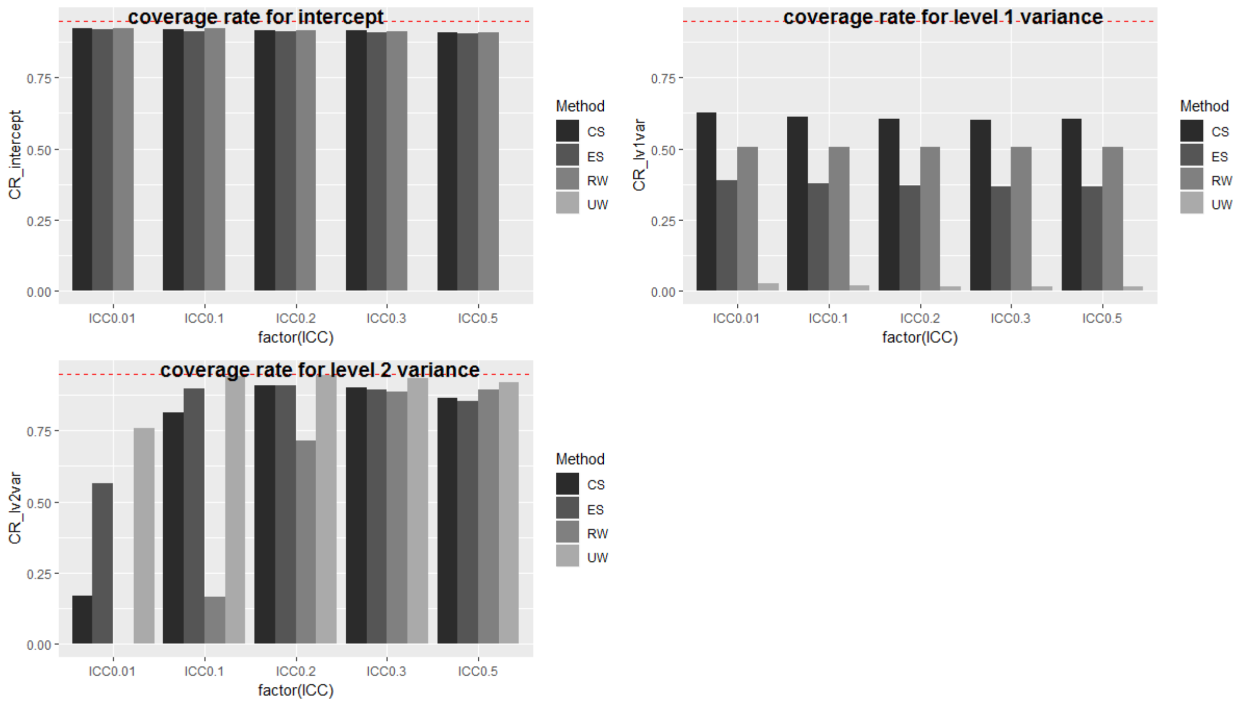

Figure 5. Coverage rate for intercept and variance components in the informative design

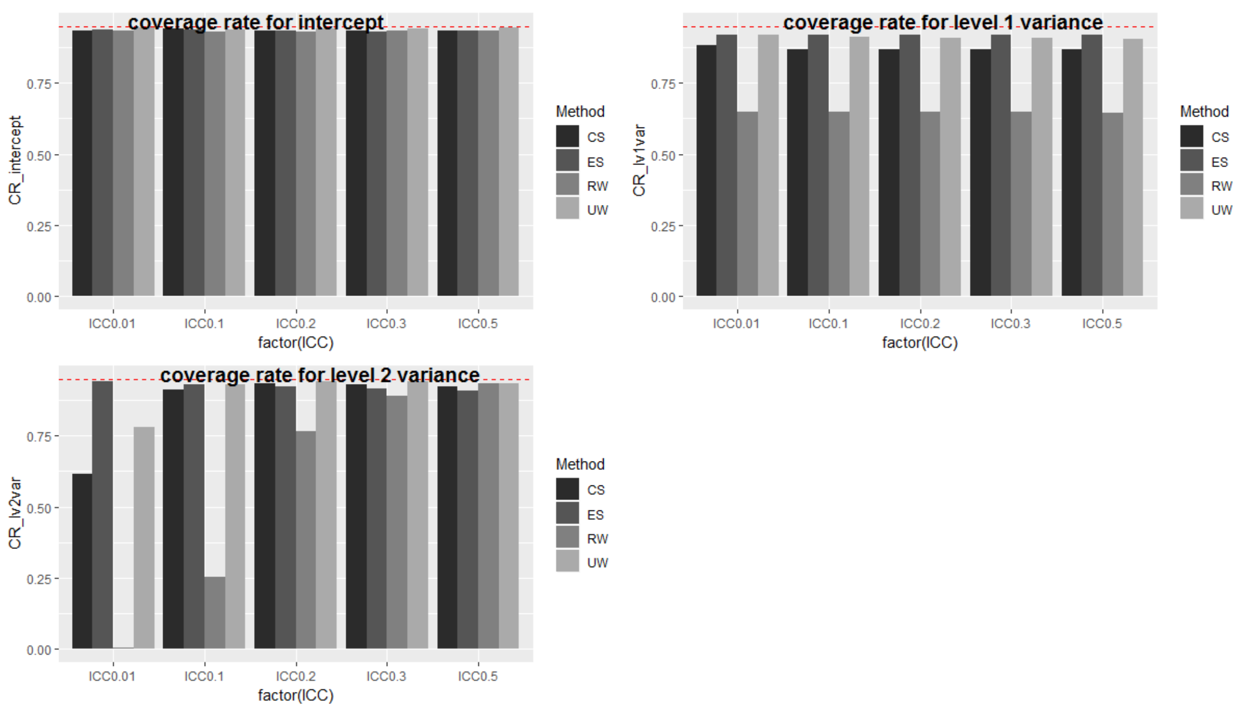

Figure 6. Coverage rate for intercept and variance components in the non-informative design