# ROBUSTNESS OF MULTILEVEL PARAMETER ESTIMATES AGAINST SMALL SAMPLE SIZES

Cora J.M. Maas[1]
*Utrecht University, The Netherlands*
Joop J. Hox
*Utrecht University, The Netherlands*

In social sciences, research questions often refer to hierarchically structured data. For instance, the achievements of pupils in classes (in schools) are studied, or the work satisfaction of employees in organizations. The main problem of studying such hierarchical systems, which have more than one level, is the dependence of the observations at the lower levels. Multilevel analyzing programs account for this dependence and in recent years these programs are widely accepted.

In this paper we will discuss the influence of different circumstances on the robustness of the parameter estimates (regressioncoefficients and the variance) in two level situations. A simulation study is used to determine the influence of small sample sizes at both level one and level two, and different variance distributions between the levels (the so called ‘intraclass correlation’).

*Key words*: Multilevel analysis, robustness of parameter estimates

## INTRODUCTION

Social science often studies systems with a hierarchical structure. For instance, the achievements of pupils in classes (in schools) are studied, or the work satisfaction of

[1] E-mail: C.Maas@fss.uu.nl and J.Hox@fss.uu.nl

employees in organizations. This hierarchical structure of the data creates problems, because the observations at the lower levels are not independent. Multilevel analysis techniques have been developed for the linear regression model, which account for this dependence (Bryk & Raudenbush, 1992; Goldstein, 1995), and specialized software is now widely available (Bryk, Raudenbush & Congdon, 1996; Rasbash & Woodhouse, 1995).

In this paper we will discuss the influence of different circumstances on the robustness of the parameter estimates in two level situations. Three conditions are varied in the simulation: (1) Number of Groups (3 conditions: 30, 50 and 100), (2) Group Size (3 conditions: 5, 30 and 50) and (3) Intraclass Correlation (3 conditions: low: 0.1, medium: 0.2 and high 0.3).

## The multilevel model

We will use a simple two level model, for data obtained from *N* individuals, nested within *J* groups, each containing $N_j$ individuals. At the individual level we specify one explanatory variable ($X_{ij}$) and at the group level also one explanatory variable ($Z_j$). The following model is specified:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{oj} + e_{ij} \qquad (1.1)$$

The u-terms $u_{0j}$ and $u_{1j}$ are residual error terms at the group level, with variance $\sigma_{00}^2$ and $\sigma_{11}^2$. The covariance between the u-terms is $\sigma_{01}$. The $e_{ij}$ is the residual error at the indivudual level, with variance $\sigma_e^2$. In the above specified model four fixed parameters have to be estimated (the γ-coefficients) and four random parameters (the variance components).

If we delete all explanatory variables of equation (1.1), we get the 'intercept-only-model, as specified in equation (1.2)

$$Y_{ij} = \gamma_{00} + u_{oj} + e_{ij} \qquad (1.2)$$

This equation is used to calculate the intraclass correlation (ICC), that is the estimated proportion of group level variance compared to the estimated total variance:

$$ICC = \sigma_{00}^2 / \sigma_{00}^2 + \sigma_e^2 \qquad (1.3)$$

## Sample sizes

In hierarchically nested data, we have two sample sizes. First the Group Size (GS), that means the number of $N_j$ indivuduals in group *J*, and second the Number of Groups (NG), that is the number of *J* groups. Both conditions have influence on the estimates of the above specified parameters. For the estimation of the fixed parameters and their standard errors a large nuber of groups appears more important than a large number of individuals per group (Van der Leeden & Busing, 1994; Mok,1995; Snijders & Bosker, 1994). The estimates of the random parameters and their standard errors at the lowest level are generally accurate. The group level variance components are generally underestimated. For accurate estimates many groups (more than 100) are needed (Busing, 1993; Van der Leeden & Busing, 1994; Afshartous, 1995). Kreft (1996) suggests a rule of thumb wich she calls the '30/30' rule. To be on the safe side, with respect to the estimates of all the parameters and their standard errors, a sample of at least 30 groups with at least 30 individuals per group is necessary.

## The intraclass correlation

Another interesting question concerns the influence of the intraclass correlation on the estimation of the parameters, although the intraclass correlations can't be influenced by the researcher, which was the case in the above discussed influence of the sample sizes. Simulation studies by Muthén, Wisnicky and Nelson (1991) and Hox and Maas (2001) in the context of multilevel structural equation modeling (SEM) suggest that the intraclass correlation also affects the accuracy of the estimates.

# METHOD

## The Simulation Model

We used the above specified simple two level model (equation 1.1) for the simulation study. The only exception is that there was no covariance specified at the second level.

## Simulation Procedure

We will look at three conditions in this simulation: (1) Number of Groups (3 conditions: 30, 50 and 100), (2) Group Size (3 conditions: 5, 30 and 50) and (3) Intraclass Correlation (3 conditions: low: 0.1, medium: 0.2 and high 0.3).

The number of groups (30-50-100) is chosen so that the highest number is almost the maximum achievable number. Only in longitudinal research or large-scale research more than 100 units at the second level are usual. In practice, 50 groups is a frequently occuring number. Although, according to Busing (1993) this is possibly too low, Kreft (1996) argued that 30 groups is the absolute minimum.

In longitudinal research, the lowest level is the number of repeated measures. A number of 5 often occurs. Also in family research, the number of 5 units at the lower level is quite normal (parents with 3 children). Therefore, the lowest group size is 5. In educational research, a group size of 30 is normal. (See also the 30/30 rule of Kreft 1996). The highest number, a group size of 50, must be sufficient.

In educational research, most ICC's are below 0.20. However, in family research, or when group chararchteristics such as sociometric status are studied, ICC's above 0.33 do occur.

There are 3x3x3=27 conditions. For each condition, we generate 1000 data sets of two variables. Both the lower level variable X, as the second level variable Z, are randomly drawn from a normal distribution. The mulilevel program MLn (Rasbach & Woodhouse, 1995) is used to specify the model and to generate the simulation data. First the intercept and the crosslevel interaction between the variabels X and Z are computed. The regression coefficients are specified as follows: 1.00 for the intercept and the medium effect of .3 for the other regression coefficients (Cohen, 1988). The random variance of the intercept at the lowest level ($\sigma_e^2$) is .5. The random variance of the intercept at the second level ($\sigma_{00}^2$) follows from the specification of the ICC and $\sigma_e^2$, given formula (1.3). Busing (1993) shows that both the pattern and the size of the $\sigma_{00}^2$ and the $\sigma_{01}^2$ (random variances of the variables of the lower level) are the same. So, the same value for $\sigma_{01}^2$ is used as for $\sigma_{00}^2$. To keep a relative simple model, the covariance between $\sigma_{00}^2$ and $\sigma_{01}^2$ is set to zero.

The foregoing sumulation decisions are summarized in table 1.

Table 1

Simulation decisions

| Assumption | Formula | Fixing |
|---|---|---|
| $X_{ij} \sim N(0,1)$ | | $X_{ij}$ |
| $Z_j \sim N(0,1)$ | | $Z_j$ |
| $\sigma_e^2 = 0.5$ | | $\sigma_e^2$ |
| ICC = (0.1, 0.2, 0.3) | ICC = $\sigma_{00}^2/(\sigma_{00}^2 + \sigma_e^2)$ | $\sigma_{00}^2$ |
| $\sigma_{00}^2 = \sigma_{01}^2$ | | $\sigma_{01}^2$ |
| NG = (30, 50, 100) | | NG |
| GS = (5, 30, 50) | | GS |

Two different estimations methods are used in the multilevel software, the 'Restricted Iterative Generalized Least Squares' (RIGLS) estimation method and the Iterative Generalized Least Squares (IGLS) estimation method. In this paper the RIGLS method is used, because this method is always as good as the IGLS method and sometimes better (Browne, 1998).

## RESULTS

### Analysis

The relative bias and the 'Mean squared error' (MSE) are commonly used measures of the accuracy of the parameter estimates. Let $\hat{\theta}$ be the estimation of the population parameter $\theta$. Then the percentage relative bias is given by: $\frac{\hat{\theta}}{\theta}$. The MSE is given by: $\sqrt{(\hat{\theta}-\theta)^2}$. In addition we also present the observed coverage of the 95% confidence interval.

**Relative bias of the parameter estimates**

The fixed parameter estimates have an overall relative bias of 0.0%, with no differences across the conditions. There are no significant effects of the specified conditions ($\alpha$ = 0.01). In table 2 the overall relative bias for the fixed effects (regression coefficients) in the various conditions are presented. The estimated mean bias in all the conditions is 1.00. The mean estimates of the parameters are thus almost perfect.

Table 2

Relative bias of fixed parameters

| Number of Groups | Group Size | ICC Low (0.1) | ICC Medium (0.2) | ICC High (0.3) |
|---|---|---|---|---|
| 30 | 5 | .998 | 1.007 | .992 |
| | 30 | 1.001 | 1.002 | 1.003 |
| | 50 | .998 | 1.000 | 1.003 |
| 50 | 5 | 1.003 | 1.000 | 1.000 |
| | 30 | 1.001 | 1.002 | 1.001 |
| | 50 | 1.000 | 1.002 | .997 |
| 100 | 5 | 1.000 | .999 | .998 |
| | 30 | 1.001 | 1.000 | 1.001 |
| | 50 | 1.001 | .999 | 1.000 |

The estimates of the random parameters at the second level have also an overall relative bias of 0.0% (no significant effects). Tabel 3 shows the overall relative bias. The mean estimates show some deviations of 0.01% and one of 0.04%. So, also the mean estimates of the random effects of the second level are almost perfect.

Table 3

Relative bias of random parameters at level 2

| Number of Groups | Group Size | ICC Low (0.1) | Medium (0.2) | High (0.3) |
|---|---|---|---|---|
| 30 | 5 | 1.037 | 1.009 | .992 |
| | 30 | 1.010 | 1.004 | .993 |
| | 50 | 1.004 | 1.005 | .994 |
| 50 | 5 | 1.008 | .997 | .994 |
| | 30 | .997 | .991 | 1.011 |
| | 50 | 1.006 | 1.010 | 1.000 |
| 100 | 5 | 1.004 | .998 | 1.006 |
| | 30 | 1.010 | .997 | 1.003 |
| | 50 | .997 | 1.000 | 1.001 |

The estimates of the random parameter of the lowest level has an overall relative bias of 0.0%. There is a significant effect of the Group Size (p = 0.005). In Table 4 the relative bias of this variance component is presented. In table 5 the overall relative bias for all the conditions are presented. The differences in the relative mean bias for the different conditions in table 4 are less than 0.01%. In table 5, the mean estimates show some deviations of 0.01%.

Concluding, we can say that the relative bias of the parameter estimates is negligible.

Table 4

Relative bias of random parameter at level 1 with respect to the number of persons per group

| Group Size | rel. bias |
|---|---|
| 5 | .997 |
| 30 | 1.000 |
| 50 | 1.000 |

Table 5

Relative bias of the random parameter at level 1

| Number of Groups | Group Size | ICC Low (0.1) | Medium (0.2) | High (0.3) |
|---|---|---|---|---|
| 30 | 5 | .990 | .998 | .998 |
| | 30 | .998 | 1.000 | .999 |
| | 50 | 1.000 | 1.002 | 1.000 |

| | | | | |
|---|---|---|---|---|
| 50 | 5 | .998 | 1.005 | .995 |
| | 30 | 1.001 | 1.001 | 1.001 |
| | 50 | .999 | 1.000 | .999 |
| 100 | 5 | .997 | .995 | .998 |
| | 30 | 1.000 | 1.000 | 1.000 |
| | 50 | 1.001 | 1.001 | 1.001 |

**Absolute bias of the parameter estimates**

The analyses of the absolute bias of the parameters show different results than the above presented relative bais.

With respect to the estimates of the fixed parameters, the overall absolute bias is 0.0%, but all the conditions, and interactions between them, have significant effects. Though, the size of the absolute bias is very little. Only the signficiant effects of the ICC are of interest, because the effects of Number of Groups and Group Size are simply a reflection of less sampling variability due to larger sample sizes.

In table 6, the absolute bias for different values of the ICC is presented. The differences between the ICC-conditions are less than 0.01. The interactions of the ICC with the Number of Groups and Group Size have also differences between conditions less than 0.01.

Table 6
Absolute bias of the ICC

| ICC | abs. bias |
|---|---|
| low | 0.029 |
| medium | 0.035 |
| high | 0.041 |

**95% confidence intervals**

For each parameter is counted how many times the confidence intervals cover the true popopulation value of the parameter. Most of the random parameters are affected by the

Number of Groups and the Size of the Group. Some of the fixed parameters are affected by the Number of Groups too. Table 7 represents the influence of the Number of Groups and table 8 the influence of the Group Size.

Table 7
Influence of the Number of Groups on the non-covarage of the 95% confidence interval

| | Number of Groups | | |
|---|---|---|---|
| | 30 | 50 | 100 |
| U0 | 0.089 | 0.074 | 0.060 |
| U1 | 0.088 | 0.072 | 0.057 |
| E0 | 0.058 | 0.056 | 0.049 |
| INT | 0.064 | 0.057 | 0.053 |
| X | 0.064 | 0.057 | 0.052 |

Table 8
Influence of the Group Size on the non-covarage of the 95% confidence interval

| | Group Size | | |
|---|---|---|---|
| | 5 | 30 | 50 |
| U1 | 0.074 | 0.075 | 0.074 |
| E0 | 0.061 | 0.051 | 0.051 |

Concluding we can state, that only when the Number of Groups is 30, the covarage of the 95% confidence interval of the variance components of the second level is really too small (almost 9% outside the interval). All the other effects are of minor importance.

**DISCUSSION**

The conditions that were varied in this simulation have very little impact on the accuracy of the estimates of the fixed parameters. Only the confidence intervals and hence the significance tests of the random parameters at the second level are not accurate when the number of groups is as small as 30. Almost 9 percent of the random parameter estimates of the second level lay outside the 95% percent confidence interval when the number of groups is 30. We conclude therfore that the 30/30 rule of Kreft (1996) only holds with respect to the fixed parameters and the random parameter of the lowest level and not with respect to the significance tests of the random parameters of the second level.

A remarkable result of this study is further that all simulations did converge. This is in contradiction with Busing (1993), who found nonconvergence especially in conditions with 10 to 50 groups and in conditions with 5 or 10 individuals in the groups. Also in contradiction with Leeden & Busing (1994) are the surprisingly accurate estimates. Are these the results of better software? Another explanation could be that these results are found because we have a model with a relatively high percentage explained variance at the lower level (about 35 percent). However, when we repeated the simulations with only about 10 percent explained variance, almost the same results were found.

A final remark we want to make is about the estimation procedure. The simulations are done with the RIGLS-estimation procedure. This procedure gives always better (or the same) results as the IGLS-estimation procedure. So the results can not be generalized to IGLS-estimations.

## References


Afshartous, D. (1995): Determination of Sample Size for Multilevel Model Design. *Paper AERA Conference*, San Francisco, 18-22 april 1995.

Browne, (1998)

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical Linear Models*. Newbury Park, CA: Sage.

Busing, F. (1993): *Distribution Characteristics of Variance Estimates in Two-level Models.* Department of Psychometrica and research Methodology, Leiden University, Leiden.

Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1996). *HLM. Hierarchical Linear and Nonlinear Modeling with the HLM/2L adn HLM/3L Programs*. Chicago: Scientific software International.

Cohen, J. (1988). *Statistical Power Analysis*. Hillsdale, New Jersey, Erlbaum.

Goldstein, H. (1995). *Multilevel Statistical Model*. London: Griffin.

Hox, Joop J. & Maas, Cora J.M. (2001). Multilevel SEM with Pseudobalanced Groups. *Structural Equation Modeling*.

Kreft, Ita, G.G. (1996). *Are Multilevel Techniques Necessary? An Overview, Including Simulation Studies*. California State University, Los Angeles.

Mok, M. (1995): 'Sample size requirements for 2-level designs in educational research'. *Multilevel Modeling Newsletter*, 7 (2), 11-15.

Muthén, B., Wisnicky, K.S. & Nelson, G. (1991). *Estimation of covariance structure models with multilevel (clustered) data*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April 3-7, 1991.

Rasbash, J. & Woodhouse, G. (1995). *MLn Command Reference*. London: Institute of Education, University of London.

Snijders, T.A.B. & Bosker, R. (1994): Modeled Variance in Two-level Models. *Journal of Educational Statistics*, 18, 273-259.

Van der Leeden, R. & Busing, F. (1994). *First Iteration versus IGLS/RIGLS Estimates in Two-level Models: a Monte Carlo Study with ML3*. Department of Psychometrica and research Methodology, Leiden University, Leiden.