J. R. Statist. Soc. A (1996)
159, *Part* 2, *pp.* 201–212

Adjusting for Measurement Error in Multilevel Analysis

By GEOFFREY WOODHOUSE†, MIN YANG, HARVEY GOLDSTEIN and JON RASBASH

Institute of Education, London, UK

[Received May 1995. Final revision October 1995]

SUMMARY

The effects of adjustment for measurement error are illustrated in a two-level analysis of an educational data set. It is shown how estimates and conclusions can vary, depending on the degree of measurement error that is assumed to exist in explanatory variables at level 1 and level 2, and in the response variable. The importance of obtaining satisfactory prior estimates of measurement error variances and covariances, and of correctly adjusting for them during analysis, is demonstrated.

Keywords: HIERARCHICAL DATA; MEASUREMENT ERROR COVARIANCE; MEASUREMENT ERROR VARIANCE; MULTILEVEL MODEL; RELIABILITY; VARIANCE COMPONENTS

1. INTRODUCTION

Most measurements in educational and other social research are subject to error, in the sense that a repetition of the measurement process does not produce an identical result. For example, measurements of cognitive outcomes in schools such as scores on standardized tests can be affected by item inconsistency, by fluctuations within individuals and by differences in the administration of the tests and in the environment of the schools and classes where the tests take place. Measurements of non-cognitive outcomes also, such as children's behaviour, self-concept and attitudes to school, can be similarly affected. It is well known that the use of such measurements in analysis, without taking measurement errors into account, can lead to mistaken causal inferences. Goldstein (1979) showed, in an analysis of social class differences in the educational attainment of children aged 11 years, how a conclusion can be reversed when a correction is introduced for measurement error. Fuller (1987) has given a comprehensive account of methods for dealing with errors of measurement in regression models but observed that few statistical studies appear to use such procedures. Plewis (1985) reviewed methods of correcting for measurement error proposed by Degracie and Fuller (1972) and by Jöreskog (1970), and explored the effects of various methods on the conclusions obtained.

All these studies are based on classical single-level regression models. Social research data, however, often have a hierarchical structure and are most efficiently analysed by means of multilevel models (Goldstein, 1995). The present paper is concerned with the effects of measurement error in continuous variables on multilevel model estimates. Methods derived by Goldstein (1986) and extended in Goldstein (1995) and in Woodhouse (1996) are applied to a two-level analysis of a longitudinal data set. We show how parameter estimates and associated conclusions

†*Address for correspondence*: Institute of Education, University of London, 20 Bedford Way, London, WC1H 0AL, UK.
E-mail: teuegmw@ioe.ac.uk

 0035–9238/96/159201

change as we change assumptions about the extent of measurement error in explanatory variables at levels 1 and 2, and in the response variable.

2. DATA SET AND MODEL

For illustration we use a sample of 1075 pupils from 48 schools. The data are derived from the Junior School Project (JSP), a longitudinal study of an age cohort of pupils who entered junior classes in September 1980 and transferred to secondary school in September 1984. (The first junior year in English and Welsh primary schools is now called year 3 of the national curriculum: pupils typically reach their eighth birthday during this year.) The schools used by the JSP were selected randomly from the 636 primary schools that were maintained by the Inner London Education Authority at the start of the project. A full account of the project is given in Mortimore *et al.* (1988).

We consider the data to have a two-level structure, with pupils (indexed by i) at level 1 and schools (indexed by j) at level 2. We use a model with the following characteristics:

(a) it contains an explanatory variable at level 1 which is subject to measurement error;
(b) it contains an aggregate explanatory variable at level 2, also subject to measurement error;
(c) it contains an explanatory variable at level 1 which is assumed to be known without error, but whose estimated coefficient may be affected by error in the other explanatory variables;
(d) it is relatively straightforward to interpret both substantively and statistically.

We have four variables:

(a) Y_{ij} is the pupil's observed reading score in year 5 (at average age 10 years), transformed by using normal scores to have a standard normal distribution;
(b) X_{1ij} is the pupil's observed reading score in year 3 (at average age 8 years), also normalized;
(c) X_{2j} is the mean normalized observed year 3 reading score for pupils in school j, as estimated from the sample;
(d) X_{3ij} is the pupil's family socioeconomic status (SES), coded 1 if the father is employed in non-manual work and 0 otherwise.

The model that we wish to estimate is based on true values and may be written

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2j} + \beta_3 x_{3ij} + u_j + e_{ij}. \tag{2.1}$$

In this model y_{ij}, x_{1ij}, x_{2j} and x_{3ij} represent true values of the variables denoted by capitals in the preceding list. In particular, x_{2j} is the mean of the true year 3 reading scores for the full cohort in school j. The u_j and e_{ij} are residuals at level 2 and level 1 respectively, assumed to be independently distributed with zero mean and constant variance:

$$\left.\begin{aligned} \operatorname{cov}(u_j, u_{j'}) = \operatorname{cov}(e_{ij}, e_{i'j}) = 0, \\ \operatorname{cov}(u_j, e_{ij}) = 0, \\ E(u_j) = E(e_{ij}) = 0, \\ \operatorname{var}(u_i) = \sigma_u^2, \\ \operatorname{var}(e_{ij}) = \sigma_e^2. \end{aligned}\right\} \quad (2.2)$$

The parameters to be estimated are β_0, β_1, β_2, β_3 (the fixed parameters) and σ_u^2 and σ_e^2 (the random parameters). Equations (2.1) and (2.2) define a two-level variance components model for the true year 5 reading score of each pupil, conditional on the true year 3 score of the pupil, the true mean year 3 score for the cohort in the school, and the pupil's true family SES. The observed individual scores and the mean year 3 scores are subject to measurement error,

$$\left.\begin{aligned} X_{1ij} = x_{1ij} + \xi_{1ij}, \\ X_{2j} = x_{2j} + \xi_{2j}, \\ Y_{ij} = y_{ij} + \eta_{ij}. \end{aligned}\right\} \quad (2.3)$$

Each pupil's SES is assumed to be observed without error:

$$X_{3ij} = x_{3ij}. \quad (2.4)$$

We assume that the errors ξ_{1ij} are distributed with zero mean and constant variance, and that they are not correlated either with one another across pupils or with the true values x_{1ij}. We make similar assumptions about the errors η_{ij}. The errors ξ_{2j} are assumed to be distributed with zero mean, and their variances will depend on the numbers of pupils in each school j, in the manner described in Section 3. They will also covary with the errors ξ_{1ij}.

Considerable further elaboration would be required to produce a fully satisfactory model for these data. For example, it is unrealistic to assume complete reliability for the SES variable, it is probable that the coefficient of x_1 varies from school to school, and other explanatory variables should be considered for inclusion. It is also a simplification to assume that each pupil's year 5 score will be affected equally by the mean year 3 score for the full cohort in the school. The purpose of the analysis, however, is to illustrate in a relatively simple context the effects of adjustment for measurement error.

3. MEASUREMENT ERROR VARIANCES AND COVARIANCES, AND RELIABILITY

The model described by equations (2.1)–(2.4) can be estimated by using the multilevel modelling package *MLn* (Rasbash and Woodhouse, 1995), which incorporates the theory described in Woodhouse (1996). Before carrying out this analysis, however, we require prior estimates of the measurement error variances of X_1 and of Y, and for each school j an estimate of the measurement error variance of X_2 and its measurement error covariance with the errors in X_1 for pupils within the school.

Ecob and Goldstein (1983) discussed the difficulty of obtaining satisfactory estimates of measurement error variances and covariances. They cast doubt on the assumptions underlying standard procedures and proposed an alternative procedure based on the use of instrumental variables. In the present paper we do not address the issue of measurement error estimation directly, but we observe that in general estimates of measurement error variances are imprecise and it is advisable to investigate the effects of varying assumptions. In the present case, no dependable prior estimates are available for the measurement error variances of Y and X_1; nor are there suitable estimates of the reliability of these variables for the population under study.

The analyses in Section 4 show the effects on the parameter estimates of adjustment for varying amounts of measurement error. These amounts are expressed in the form of assumed reliabilities for Y and X_1. We now derive expressions which will enable us to convert these assumptions into the measurement error variances and covariances needed for the parameter estimation.

If we assume that the measurement error variance of X_1 is constant for all pupils, with a known value $\mathrm{var}(\xi_1)$, it is straightforward to estimate for each school the variance of the error in X_2 and its covariance with the errors in X_1 for pupils in the school.

Consider a given school j. Let N_j be the size of the cohort in school j, and write n_j for the size of the sample. We have, by definition,

$$X_{2j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{1ij}. \tag{3.1}$$

The measurement error ξ_{2j} in X_{2j} is given by

$$\begin{aligned} \xi_{2j} = X_{2j} - x_{2j} &= (X_{2j} - \bar{x}_{1.j}) + (\bar{x}_{1.j} - x_{2j}) \\ &= \frac{1}{n_j} \sum_{i=1}^{n_j} \xi_{1ij} + \bar{x}_{1.j} - x_{2j}, \end{aligned} \tag{3.2}$$

where $\bar{x}_{1.j}$ is the mean of the true scores for the pupils sampled, and $E(\bar{x}_{1.j}) = x_{2j}$.

For school j, x_{2j}, though unknown, is fixed. The sample of size n_j has been selected, we assume at random without replacement, from the known cohort of size N_j. Therefore

$$\begin{aligned} \mathrm{var}(\xi_{2j}) &= \frac{1}{n_j} \mathrm{var}(\xi_1) + \mathrm{var}(\bar{x}_{1.j}) \\ &= \frac{1}{n_j} \mathrm{var}(\xi_1) + \frac{1}{n_j} \sigma_{\mathrm{w}}^2(x_{1ij}) \frac{N_j - n_j}{N_j - 1}, \end{aligned} \tag{3.3}$$

where $\sigma_{\mathrm{w}}^2(x_{1ij})$ is the variance of the true scores x_1 within the cohort for school j.

Since measurement errors are assumed to be uncorrelated across students, the measurement error covariance between X_1 and X_2 for pupil a in school j is

$$\begin{aligned}\operatorname{cov}(\xi_{1aj}, \xi_{2j}) &= \operatorname{cov}\left(\xi_{1aj}, \frac{1}{n_j} \sum_{i=1}^{n_j} (x_{1ij} + \xi_{1ij}) - \frac{1}{N_j} \sum_{i=1}^{N_j} x_{1ij}\right) \\ &= \operatorname{cov}\left(\xi_{1aj}, \frac{1}{n_j} \sum_{i=1}^{n_j} \xi_{1ij}\right) \\ &= \operatorname{cov}\left(\xi_{1aj}, \frac{\xi_{1aj}}{n_j}\right) \\ &= \frac{1}{n_j} \operatorname{var}(\xi_1).\end{aligned} \tag{3.4}$$

These results enable us to estimate model (2.1)–(2.4), provided that we have estimates of, or make assumptions about,

(a) $\sigma_w^2(x_{1ij})$, the within-school variance of the true year 3 scores, for each school,
(b) $\operatorname{var}(\xi_1)$, the measurement error variance of the observed year 3 scores, and
(c) σ_η^2, the measurement error variance of the observed year 5 scores.

We shall assume that the within-school variance $\sigma_w^2(x_{1ij})$ of the true year 3 scores is the same for all schools, with value $\sigma_w^2(x_1)$. As we have already assumed that the measurement errors in the observed scores are uncorrelated with the true scores and that their variance is constant, it follows that the within-school variance of the observed year 3 scores also is constant, say $\sigma_w^2(X_1)$. A reasonable estimate of this is the level 1 variance obtained when fitting X_1 to its mean in a two-level variance components model.

We do not have a prior estimate of $\operatorname{var}(\xi_1)$, and one possible way forward is to define the *level 1 reliability* R_1 of X_1 as

$$\begin{aligned}R_1 &= \sigma_w^2(x_1)/\sigma_w^2(X_1) \\ &= 1 - \operatorname{var}(\xi_1)/\sigma_w^2(X_1).\end{aligned} \tag{3.5}$$

Now, given an estimate $\hat{\sigma}_w^2(X_1)$ for the within-school variance of X_1, and an assumed value of R_1, the level 1 measurement error variance of X_1 is estimated as

$$\widehat{\operatorname{var}}(\xi_1) = (1 - R_1)\, \hat{\sigma}_w^2(X_1). \tag{3.6}$$

With these assumptions, the level 2 measurement error variance of X_{2j} is estimated as

$$\begin{aligned}\widehat{\operatorname{var}}(\xi_{2j}) &= \frac{1}{n_j}\, \widehat{\operatorname{var}}(\xi_1) + \frac{1}{n_j}\, \hat{\sigma}_w^2(x_1)\, \frac{N_j - n_j}{N_j - 1} \\ &= \frac{1}{n_j}\, (1 - R_1)\, \hat{\sigma}_w^2(X_1) + \frac{R_1}{n_j}\, \hat{\sigma}_w^2(X_1)\, \frac{N_j - n_j}{N_j - 1} \\ &= \frac{\hat{\sigma}_w^2(X_1)}{n_j} \left(1 - R_1\, \frac{n_j - 1}{N_j - 1}\right),\end{aligned} \tag{3.7}$$

and the covariance with errors in X_1 as

$$\widehat{\text{cov}}(\xi_{1ij}, \xi_{2j}) = \frac{1}{n_j}\, \widehat{\text{var}}(\xi_1) = \frac{1}{n_j}\, (1 - R_1)\, \hat{\sigma}^2_{\text{w}}(X_1). \tag{3.8}$$

Similarly, if we assume that the within-school variance of the true year 5 scores is constant, it follows that the within-school variance of the observed response also is constant. Given an estimate $\hat{\sigma}^2_{\text{w}}(Y)$ for this variance, and an assumed value R_Y of the *response reliability*, we have the following estimate for the measurement error variance of Y:

$$\hat{\sigma}^2_\eta = (1 - R_Y)\, \hat{\sigma}^2_{\text{w}}(Y). \tag{3.9}$$

Note that the adjustment method described in Woodhouse (1996) uses measurement error variances and covariances, not reliabilities. Our assumption here of constant reliabilities across schools for X_1 and Y allows us to express in a familiar way different assumptions about the degree of measurement error in the data. This use of reliabilities is not essential, and in particular we make no use of the reliability of X_2 in the population of schools. (See Raudenbush *et al.* (1991) for a further discussion of this issue.)

Longford (1993) described a measurement error model where a vector of indicator or manifest variables S is a linear function of a set of true but unobserved variables x. At its most general this is a latent variable model and in the special case of a single unobserved variable it is the usual congeneric test score model. Longford specialized the model to the case where the 'loadings' are known, and in particular he considered the assumption of 'exchangeability' where the loadings are assumed to be equal. For these cases he derived maximum likelihood estimators. Longford's model is therefore different from the present case where we assume prior knowledge of the measurement error variance rather than specifying an implicit model which allows the measurement error variance to be estimated.

4. ANALYSIS AND DISCUSSION

Four analyses of the model in Section 2 were conducted, illustrating the effects of progressively more complete adjustment:

(a) analysis A, adjustment for measurement error in X_1 only, i.e. no adjustment for measurement error at level 2 or in the response;
(b) analysis B, adjustment for measurement errors in X_1 and X_2 but not for covariances between these errors nor for measurement error in the response;
(c) analysis C, adjustment for measurement errors in X_1 and X_2 and for covariances between them, but no adjustment for measurement error in the response;
(d) analysis D, adjustment for measurement errors in X_1 and X_2 and for covariances between them, and for measurement error in the response.

Unadjusted results were provided by analysis A with $R_1 = 1.0$. In all analyses X_3 was assumed to have been measured without error. Of the four analyses, clearly analysis D is to be preferred in practice. The other three show the extent to which it is possible to be misled by incomplete adjustment for errors in the variables.

TABLE 1
Sample and cohort sizes for each school (indexed by j)†

j	n_j	N_j	n_j/N_j	j	n_j	N_j	n_j/N_j	j	n_j	N_j	n_jN_j
1	6	20	0.30	17	16	38	0.42	33	24	47	0.51
2	7	25	0.28	18	16	59	0.27	34	25	31	0.81
3	7	43	0.16	19	18	28	0.64	35	25	41	0.61
4	11	22	0.50	20	18	29	0.62	36	26	53	0.49
5	12	22	0.55	21	18	37	0.49	37	28	55	0.51
6	12	29	0.41	22	18	51	0.35	38	29	53	0.55
7	14	29	0.48	23	18	53	0.34	39	29	58	0.50
8	14	29	0.48	24	19	31	0.61	40	32	54	0.59
9	14	32	0.44	25	19	32	0.59	41	32	82	0.39
10	14	33	0.42	26	21	25	0.84	42	33	65	0.51
11	15	26	0.58	27	21	37	0.57	43	35	49	0.71
12	15	31	0.48	28	21	45	0.47	44	38	53	0.72
13	15	39	0.38	29	21	51	0.41	45	39	52	0.75
14	15	43	0.35	30	23	34	0.68	46	44	61	0.72
15	15	44	0.34	31	23	40	0.58	47	52	92	0.57
16	16	31	0.52	32	24	43	0.56	48	68	116	0.59

†n_j is the number of pupils sampled; N_j is the cohort size.

Table 1 shows for each school j the number of pupils in the sample, the total number in the age cohort within the school and the sampling fraction. Equations (3.6)–(3.8) provide estimates for the measurement error variances of X_1 and X_2, and for the covariance between errors in X_2 and in X_1, in terms of the level 1 reliability R_1 of X_1, the estimated within-school variance $\hat{\sigma}^2_w(X_1)$, assumed constant, and the sample and cohort sizes in each school. In analyses A–C, we allow R_1 to assume values ranging from 1.0 down to 0.7. The common within-school variance of X_1 was estimated from a simple two-level variance components model fitting X_1 to its mean and was found to be 0.89.

A pooled estimate for the within-school variance of Y was obtained in a similar way from a two-level variance components model fitting Y to its mean and was found to be 0.86. Equation (3.9) was used in analysis D to provide estimates of the response measurement error variance σ^2_η for different assumptions about the response reliability R_Y.

4.1. *Parameters Essentially Unaffected by Measurement Error at Level 2*

Estimates of these parameters are summarized in Table 2. Consider first the parameter β_1. For two pupils in the same school and with the same SES, β_1 is the predicted difference in their normalized scores for year 5, per point of difference in their true normalized year 3 scores. With no adjustment for measurement error (first row of Table 2), the model predicts that the pupil scoring higher in year 3 will score higher in year 5 also, by an amount 0.77 per point on the year 3 scale. When adjustment is made at level 1, $\hat{\beta}_1$ increases as the assumed level 1 reliability R_1 decreases, so that, when $R_1 = 0.8$, $\hat{\beta}_1$ has increased by a factor 1.27. Its standard error has increased by a factor 1.41. These disattenuation factors become 1.46 and 1.74 respectively, when $R_1 = 0.7$. Thus the estimate of the coefficient of x_1 is approximately, though not exactly, in inverse proportion to the assumed level 1

TABLE 2

Adjusted estimates of parameters which were essentially unaffected by measurement error at level 2†

Reliability R_1	*Intercept* $\hat{\beta}_0$	*Year 3 score* $\hat{\beta}_1$	*SES* $\hat{\beta}_3$	*Level 1 variance* $\hat{\sigma}_e^2$
1.0	−0.037 (0.04)	0.77 (0.02)	0.17 (0.04)	0.30 (0.01)
0.9	−0.026 (0.04)	0.86 (0.02)	0.13 (0.04)	0.24 (0.01)
0.8	−0.013 (0.04)	0.97 (0.03)	0.08 (0.05)	0.17 (0.01)
0.7	0.005 (0.04)	1.12 (0.03)	0.00 (0.05)	0.07 (0.02)

†Standard errors are given in parentheses.

reliability of X_1, whereas the precision of the estimate decreases slightly as this reliability decreases. These effects are similar to those for simple regression in the presence of measurement error. The disattenuation of $\hat{\beta}_1$ is illustrated in Fig. 1(a). Further adjustment for measurement error at level 2 (as in analyses B and C) has negligible further effect on the estimate of this parameter, which remains statistically highly significant throughout.

The parameter β_3 is the predicted difference in outcome score between two pupils in the same school and with the same true score in year 3, where one pupil's father is in non-manual employment and the other's is not. With no adjustment for measurement error in X_1, the benefit to the pupil with a non-manual background is estimated to be 0.17 points on the outcome scale and to be statistically highly significant. When $R_1 = 0.8$ and adjustment is made at level 1, that estimate is attenuated by a factor 0.44, with an accompanying increase in its standard error. With the usual distributional assumptions, the effect of SES on outcome score ceases to be statistically significant at the 5% error level for $R_1 < 0.82$. As with β_1, further adjustment for measurement error at level 2 has negligible further effect on the estimate of β_3. The attenuation of $\hat{\beta}_3$ is illustrated in Fig. 1(b).

The other parameter (apart from the intercept) whose estimates do not vary noticeably between the three analyses A, B and C is the residual variance at level 1, σ_e^2. When adjustment is made on the assumption that $R_1 = 0.8$, the estimate is reduced by 44% of its unadjusted value. This is the variance in Y explained by the assumed measurement error variance of X_1 in this case. The attenuation increases to

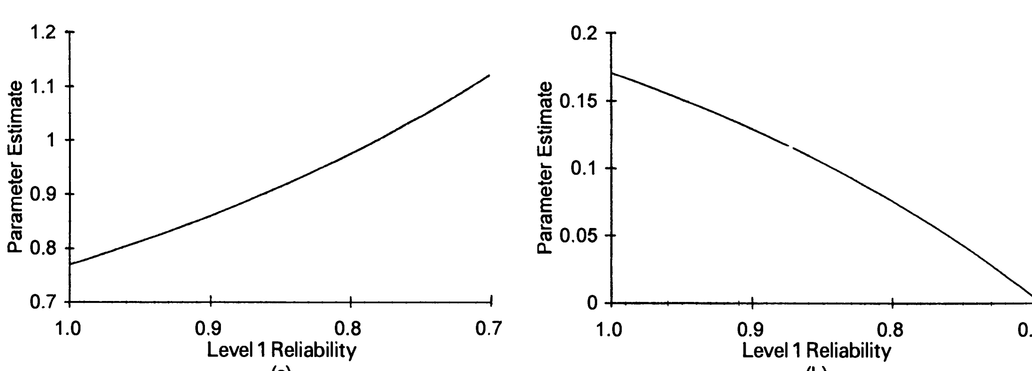

Fig. 1. Adjusted estimates of (a) β_1, the coefficient of individual year 3 score, and (b) β_3 the coefficient of SES, as functions of the level 1 reliability R_1

76% as R_1 decreases to 0.7, and $\hat{\sigma}_e^2$ vanishes for a sufficiently low assumed value of R_1 (approximately 0.65 for this data set).

4.2. *Effect of Adjustment for Measurement Error at Level 2*

The estimation of β_2 and, to a lesser extent, of the level 2 residual variance σ_u^2 is affected by measurement error at level 2 (Table 3). The parameter β_2 predicts the difference between pupils' outcomes in different schools which is attributable to the difference between the true mean year 3 scores in the schools. Its negative sign indicates that, according to the model in Section 2, pupils with a given year 3 score and SES obtain better scores on average in year 5 the lower the mean year 3 score is in their school. Thus, pupils with low year 3 scores are predicted to do better where they are closer to the school mean and pupils with high year 3 scores to do better the further they are from the school mean. Further analysis, incorporating interaction terms for example, would be needed to test such a hypothesis, but this is beyond the scope of the present paper.

Assuming normality, the effect of the school mean year 3 score is estimated to be non-significant when there is no adjustment for measurement error but increases markedly in size as the assumed level 1 reliability R_1 decreases. In analysis A, which adjusts at level 1 only, $\hat{\beta}_2$ is estimated to be statistically significant ($p<0.05$) for $R_1<0.80$.

Analysis B adjusts for the effect of the measurement error variances of X_{2j} calculated from equation (3.7) but ignores the covariances between the measurement errors in X_{2j} and X_{1ij}. With this adjustment the estimates $\hat{\beta}_2$ are further disattenuated and are estimated to be statistically significant ($p<0.05$) when $R_1<0.79$.

In analysis C the level 2 measurement error covariance matrix for each school includes the covariance between the measurement errors in X_{1ij} and X_{2j}. All such

TABLE 3
Adjusted estimates of parameters which were affected by measurement error at level 2†

Reliability R_1	*Estimates for the following adjustments*‡:		
	A	*B*	*C*
$\hat{\beta}_2$: *mean year 3 score*			
1.0	0.00 (0.10)	0.01 (0.12)	0.01 (0.12)
0.9	−0.08 (0.10)	−0.10 (0.13)	−0.07 (0.12)
0.8	−0.19 (0.10)	−0.23 (0.13)	−0.17 (0.13)
0.7	−0.33 (0.10)	−0.41 (0.14)	−0.30 (0.13)
$\hat{\sigma}_u^2$: *level 2 variance*			
1.0	0.059 (0.015)	0.059 (0.015)	0.059 (0.015)
0.9	0.060 (0.016)	0.059 (0.016)	0.060 (0.016)
0.8	0.059 (0.016)	0.058 (0.016)	0.061 (0.016)
0.7	0.056 (0.016)	0.051 (0.016)	0.063 (0.016)

†Standard errors are given in parentheses;
‡A, adjustment at level 1 only; B, adjustment at levels 1 and 2, ignoring measurement error correlation; C, adjustment at levels 1 and 2, including measurement error correlation.

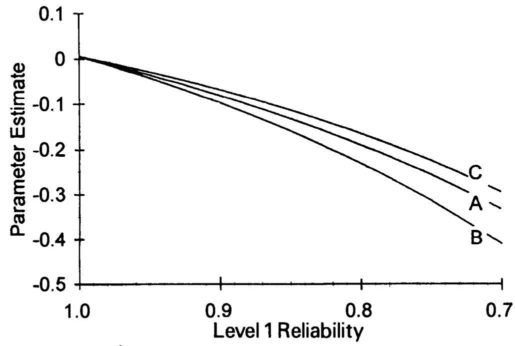

Fig. 2. Adjusted estimates of β_2, the coefficient of mean year 3 score, as functions of the level 1 reliability R_1: A, adjustment at level 1 only; B, incomplete adjustment at level 2; C, full adjustment at level 2

covariances are positive. Because the coefficients β_1 and β_2 are of opposite sign, the effect of adjusting for these positive covariances is to reduce the size of the estimates $\hat{\beta}_2$. The standard errors also are reduced but by very small amounts, so that the precision of the estimates is less than in either of the other analyses. In analysis C, $\hat{\beta}_2$ is estimated to be statistically significant ($p < 0.05$) when $R_1 < 0.73$. The disattenuation of $\hat{\beta}_2$ is illustrated in Fig. 2. If β_1 and β_2 were of like sign, the effect of adjusting for positive covariances between the measurement errors in X_{1ij} and X_{2j} would be to increase further the size of $\hat{\beta}_2$.

The estimates of level 2 residual variance σ_u^2, although somewhat different for analyses A–C, nevertheless change relatively little as a result of adjustment compared with those of σ_e^2. The estimate of the intraschool correlation therefore increases as measurement error variances increase (Fig. 3).

4.3. *Effect of Measurement Error in Response Variable*

For analysis D, estimates of the measurement error variance σ_η^2 of the response were obtained from equation (3.9), based on the assumption of equality of R_1 and the

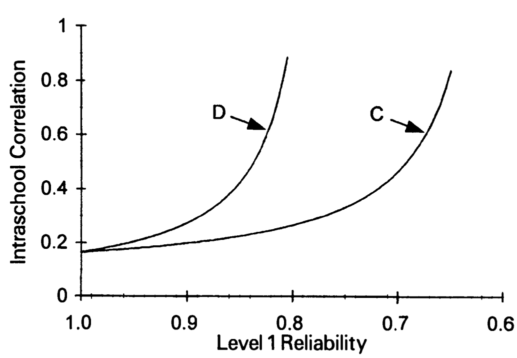

Fig. 3. Adjusted intraschool correlation as a function of level 1 reliability R_1 and response reliability R_Y: C, no adjustment for measurement error in the response; D, adjustment assuming $R_Y = R_1$

response reliability R_Y, and allowing this common value to decrease. Given R_Y, $\hat{\sigma}^2_\eta$ is constant and in particular independent of the explanatory variables in any model and of their assumed measurement error variances and covariances. Clearly, $\hat{\sigma}^2_\eta$ cannot exceed the residual level 1 variance $\hat{\sigma}^2_e$ that would be left unexplained by a model omitting $\hat{\sigma}^2_\eta$. This imposes limits on the values that R_Y and R_1 can take.

A full adjustment was made for measurement errors in Y, X_1 and X_2, and as before X_3 was assumed to have been measured without error. The lowest common value of R_1 and R_Y for which $\hat{\sigma}^2_e$ remained positive was 0.805. The estimates $\hat{\sigma}^2_e$ from analysis C were reduced, as expected, by approximately the value of $\hat{\sigma}^2_\eta$ in each case. Changes in the other parameters and their standard errors resulting from the additional adjustment for measurement error in Y were negligible.

Fig. 3 shows how the intraschool correlation depends on assumptions about R_1 and R_Y. The intraschool correlation, or more generally the intra-unit correlation, is an important statistic in multilevel analysis. In the present case a value above 0.25 would suggest a substantial school level effect not explained by the model. On the assumption of a reliability of 0.90 for Y and for X_1, we obtain a value 0.27 for the intraschool correlation. This value increases to 0.42 for $R_Y = R_1 = 0.85$. Without adjustment for measurement error in the response variable (as, for example, in analysis C) the intraschool correlation may be seriously underestimated.

5. CONCLUSIONS

Methods of adjusting for measurement error in single-level models and the effects of such adjustment are already well documented (see, for example, Fuller (1987)). In single-level regression analysis adjustment for measurement error in explanatory variables tends to increase the magnitude of the associated coefficients: our analysis has shown for this data set a similar effect at each level of a two-level model. As in the single-level case, adjustment for measurement error in a multilevel model can have substantial effects on other parameter estimates also. In this illustrative analysis the standard errors of all parameters have been found to increase when an adjustment is made.

The level 1 residual variance decreases markedly when adjustment is made for measurement error in an explanatory variable at level 1. This produces a substantial effect on the intraschool correlation, which is further increased when adjustment is made for measurement error in the response variable.

Adjustment for measurement error at level 2 has been found to be important chiefly in the estimation of parameters associated with level 2. If a level 1 variable and a level 2 aggregate of it are both present in the model, there will be measurement error variances and covariances all of which should be adjusted for. Incomplete adjustment leads to biased estimates of the parameters associated with level 2, and of their standard errors, the direction of the bias depending on the signs of the coefficients and of the covariances.

Analyses of school outcomes often use school level data based on a sample. Examples include the proportion of pupils who are eligible for free meals, the proportion of girls in the school, the proportion of pupils from specific ethnic backgrounds etc. In household surveys where we wish to model the hierarchical population structure, typically only a small percentage of households in an area are sampled, and we may wish to use area level variables based on aggregating the sample household characteristics. Thus an aggregate variable potentially includes error from two sources: error in the variable being aggregated and sampling error. The latter may be substantial, even if the former is absent. For adjustment in a multilevel model these two contributions to error in the aggregate variable may be combined as illustrated in Section 3, provided that the errors are random.

The model analysed in this paper is the simplest model capable of illustrating the effects described. Similar simple models applied to simulated data suggest that the adjusted results are indeed an improvement on the unadjusted results. Further work is needed, and it is of both practical and theoretical importance to explore the effects of adjustment for measurement error on multilevel models with random coefficients.

Finally, it is important in any study to obtain suitable estimates of measurement error variances and covariances, and in particular to investigate the effect of varying the estimates, for example, in accordance with any uncertainty intervals that may be available. Different assumptions about measurement error variances and covariances can lead to substantially different conclusions.

ACKNOWLEDGEMENTS

We are grateful to the Editors, and to two referees, whose comments have helped us to clarify the exposition.

Support for this study was given by the UK Economic and Social Research Council under grant H519255033, as part of the Analysis of Large and Complex Datasets Programme.

REFERENCES

Degracie, J. S. and Fuller, W. A. (1972) Estimation of the slopes and analysis of covariance when the concomitant variable is measured with error. *J. Am. Statist. Ass.*, **67**, 930–937.

Ecob, R. and Goldstein, H. (1983) Instrumental variable methods for the estimation of test score reliability. *J. Educ. Statist.*, **8**, 223–241.

Fuller, W. A. (1987) *Measurement Error Models*. New York: Wiley.

Goldstein, H. (1979) Some models for analysing longitudinal data on educational attainment (with discussion). *J. R. Statist. Soc.* A, **142**, 407–442.

——(1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, **73**, 43–56.

——(1995) *Multilevel Statistical Models*. London: Arnold.

Jöreskog, K. G. (1970) A general method for analysis of covariance structures. *Biometrika*, **57**, 239–251.

Longford, N. T. (1993) Regression analysis of multilevel data with measurement error. *Br. J. Math. Statist. Psychol.*, **46**, 301–311.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1988) *School Matters: the Junior Years*. Wells: Open Books.

Plewis, I. (1985) *Analysing Change: Measurement and Explanation using Longitudinal Data*. New York: Wiley.

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

Raudenbush, S. W., Rowan, B. and Kang, S. J. (1991) A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. High-School data. *J. Educ. Statist.*, **16**, 295–330.

Woodhouse, G. (1996) *Errors in Variables in Multilevel Models*. London: Institute of Education.