Adjusting for measurement error in longitudinal multilevel models

Ferrão, Maria Eugénia
University of Beira Interior (*) & CEMAPRE
(*) Rua Marquês d' Ávila e Bolama
6200-001 Covilhã, Portugal
E-mail: meferrao@ubi.pt

Measurement error is found in most of the variables used in social, behavioral and health sciences. Statistical analyses tend to omit it, and ignoring the measurement error can lead to biased estimates.

Statistical methods based on the MCMC algorithm were recently proposed to adjust for measurement errors in multilevel models with normally distributed predictor and response variables. The estimation of multilevel model parameters involving polynomial terms with errors-in-variables calls for a nonlinear framework.

We will present an extension of that MCMC algorithm involving a quadratic term. Value-added models applied to Portuguese data are used to illustrate the method. The survey design is longitudinal and consists of three waves - 2005/6, 2006/7 and 2007/8. Data were collected at the beginning and at the end of each academic year.

Introduction

Measurement error is found in most of the variables used in social, behavioral and health sciences. Ignoring measurement error and misclassification in the predictor variables of statistical models typically leads to biased parameter estimates and standard errors. Thus, a loss of power in detecting the impact of explanatory variables on the response is among the consequences. Despite the increasing efforts to encourage researchers to be more critical as regards measuring, and to adopt approaches to reduce such errors and their implications, most research papers ignore these, in part because the lack of available software that allows for the adjustment. A large statistical literature on the modelling of such errors exists, mostly dealing with the case of single level models (Fuller 2006, Carroll et al. 2006). However, the effect of measurement error in multilevel models has been far less explored. Woodhouse et al. (1996) presented an approach based upon moment type estimators in multilevel models but it does not apply to the case of random coefficients, where those coefficients are related to explanatory variables containing measurement errors. Browne et al. (2001) developed an algorithm using Markov Chain Monte Carlo estimation to deal with that case under the following assumptions: measurement errors are independent across explanatory variables; the measurement errors variances are assumed to be known; the unknown true values are assumed to have Normal distributions; the response variable is continuous. Goldstein et al. (2008) extend this work by allowing for covariances between measurement errors. The problem of categorical explanatory variables with misclassification is also addressed by those authors.

The present paper is concerned with the effects of measurement error on a value added model specified as a function of several prior year scores with quadratic terms. All of them are measured with error.

Value-added model and longitudinal data

Numerous studies in school effectiveness have aimed at identifying the factors that explain differences across schools and at assessing the magnitude, consistency and stability of school contribution to student outcomes. It is now generally agreed that performance comparisons among schools should be based on the value added by a school to the learning and development of each and every student. Thus, the family of statistical models that are employed to make inferences about the effectiveness of educational units, usually schools and/or teachers is known as value added models (Braun & Wainer, 2007). The evidence that is possible to estimate and compare the performance of institutions has had governmental implications on educational policies in several countries since there has been increasing interest in the accountability of public institutions. The measurement of such school or teacher effectiveness is crucial to the legitimacy of any school or teacher-based accountability system, particularly when it is used for high-stakes purposes. The debate has been focused on the theoretical but mainly methodological models and procedures to properly quantify the value added (e.g. Goldstein 1997, Raudenbush & Willms 1995, Goldstein and Spiegelhalter 1996, Ladd & Walsh 2002, Fielding et al. 2003, Ballou et al. 2004, Leckie & Goldstein 2009, Ferrão & Goldstein 2009). The research on the impact of measurement error on value added estimates has not been extensive. Ladd & Walsh (2002) use prior grade test scores as instrumental variable to adjust generally for measurement errors. However, additional requirements must be set because the prior grade test scores are also measured with error.

The statistical model considered here is a multilevel model for student's performance in mathematics, specified as a function of several prior grade scores with quadratic terms, all measured with error. The student's socioeconomic level is included as controlling variable.

For illustration we use a sample of 338 students (208 enrolled in primary education and 130 enrolled in lower secondary education) from 40 classes/schools. The data are derived from 3EM (Eficácia Escolar no Ensino da Matemática [School Effectiveness in Mathematics]) project[1]. The survey design is longitudinal.

The population is defined by students enrolled in compulsory education in the region of Cova da Beira, a NUT III Portuguese region. Data were collected at the beginning and at the end of academic years 2005/6, 2006/7 and 2007/8. Two cohorts of students were considered. In 2005/6 the 1st, 3rd, 5th, 7th and 8th grade students were involved. They were monitored in the 2nd, 4th, 6th, 8th and 9th grades, respectively, and a new cohort at the 1st, 3rd, 5th, and 7th grades was surveyed. In 2007/8 these students were monitored again. The random sample is representative at the level of county and NUT III region (Vicente, 2007). In 2005/6 the number of students involved from all grades was 1477 at the beginning and 1435 at the end. In 2006/7 total numbers were 3044 and 2947, respectively, and in 2007/8 the number of students was 2427 and 2370.

The Portuguese Council for Data Protection gave permission to run the survey, but conditional on parents' agreement. The initial sample was oversampled in order to take account of parents' non-agreement and dropout or attrition, which is a known problem in longitudinal studies. The analyses of data collected over the first year show that the observed sample is still representative of the target population (Vicente 2007).

For the purpose of this paper we consider 3 years longitudinal data from two group of students; those students who, in 2005/6 started the 1st grade and those enrolled at the 7th grade, are followed up to their 3rd grade (208 students) and 9th grade (130 students), respectively. The hierarchical structure is students at level 1 (indexed by i) nested in classes at level 2 (indexed by j). We use the following notation in describing the models. Let $y_{t(ij)}$ represent the mathematics achievement score (Ferrão et al. 2006) for the student i in school j at the end of grade t (t=3,9); let $p_{l(ij)}$ denote math prior score of student ij at the grade l (l=t, t-1, t-2,…) (P_3, P_2, P_1 if primary education model; P_9, P_8, P_7 if lower secondary education model). The pupil's family socioeconomic status is represented by the level of education of parents (or responsible) and it is denoted by D.

[1] The research was made possible thanks to funding by the Portuguese Ministry of Science Technology and H igher Education (2004-2007) and by the Calouste Gulbenkian Foundation (2006-2008).

For example, for primary education the model is specified by the following equation:

$$y_{3(ij)} = \beta_0 + (\beta_1 + u_{1j})p_{1(ij)} + (\beta_2 + u_{2j})p_{2(ij)} + (\beta_3 + u_{3j})p_{3(ij)} + (\beta_4 + u_{4j})p^2_{3(ij)} + \beta_5 d_{(ij)} + \\ + u_{0j} + e_{ij} \quad (1)$$

where e_{ij} is the random component or error term at level one and u_{0j}, u_{1j}, u_{2j}, u_{3j}, u_{4j} are random components at level two,

$$e_{ij} \sim N(0, \sigma^2_e),$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \\ u_{4j} \end{bmatrix} \sim MVN(0, \Omega_5), \quad \Omega_5 = \begin{bmatrix} \sigma^2_{u0} & & & & \\ \hline \sigma_{u01} & \sigma^2_{u1} & & & \\ \sigma_{u02} & \sigma_{u12} & \sigma^2_{u2} & & \\ \sigma_{u03} & \sigma_{u13} & \sigma_{u23} & \sigma^2_{u3} & \\ \sigma_{u04} & \sigma_{u14} & \sigma_{u24} & \sigma_{u34} & \sigma^2_{u4} \end{bmatrix}$$

$$COV(e_{ij}, u_{0j}) = COV(e_{ij}, u_{1j}) = COV(e_{ij}, u_{2j}) = COV(e_{ij}, u_{3j}) = COV(e_{ij}, u_{4j}) = 0.$$

Level 1 component is assumed to be independently distributed, according to a Normal distribution, and level 2 components are assumed to follow a multivariate Normal distribution with zero mean and covariance matrix Ω_5 . As we will show, most elements of this matrix are statistically equal to zero.

The parameters to be estimated are $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$, the fixed parameters, and σ^2_e, Ω_5, the random parameters. Thus, equation (1) define a two-level random component model for the true grade 3 mathematics score of each student (Y_t), conditional on the true scores, obtained per each student at the end of grade 1 (P_1), at the end of grade 2 (P_2), and at the beginning of grade 3 (P_3), and also conditional on the demographic variable parents' education which is proxy for student's socioeconomic status (D). Response variable Y_t and explanatory variables P_l, D are measured with error, so that we only observe a surrogate Y_t*, P_l*, D* respectively. According to the assumption that we are able to obtain independent replications of P_l, for example, P_{l1}*, P_{l2}*,…, P_{lk}*, here we will assume a functional model (Fuller 2006, p.2) such as the classical measurement model:

$$p^*_{1(ij)k} = p_{1(ij)} + m_{1(ij)k}$$

$$p^*_{2(ij)k} = p_{2(ij)} + m_{2(ij)k}$$

$$p^*_{3(ij)k} = p_{3(ij)} + m_{3(ij)k}$$

$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{m1} & 0 & 0 \\ 0 & \sigma^2_{m2} & 0 \\ 0 & 0 & \sigma^2_{m3} \end{bmatrix} \right)$$

where $[m1 \quad m2 \quad m3]^T$ is the vector containing the errors of measurement, assuming that the error is non-systematic, homoscedastic and $corr(P_l, m_{lk}) = 0 \ (l = 1,2,3)$, that is the true values and error measurement are not correlated. The assumption is that each true value $p_{l(ij)} \sim (\theta, \emptyset^2_l)$. Errors of measurement are assumed to be uncorrelated as the covariance matrix is set to be diagonal. Further research to allow off diagonal elements to be different from zero is in progress.

For lower secondary education the model fitted is similar to the one described above. As we shall verify for both models, random components related to the slopes are not statistically different from zero. Thus, each model reduces to a variance component model.

If the model is fitted with Y^*, P^*, and D^* instead of Y, P and D the resulting estimates will be biased.

MCMC estimation

The algorithm described in Goldstein (2003, p.217) with the extension proposed by Goldstein et al. (2008) is used to estimate model parameters, considering also the classical measurement model for each explanatory variable as explained before. A small change in the algorithm was made for estimating the coefficient related to the quadratic term of the model.

As described in the literature (e.g. Goldstein, 2003; p.56), the MCMC estimation proceeds, at each iteration, by considering each parameter in turn and generating a random sample from the distribution of that parameter assuming the current values of the remaining parameters. This procedure generates the conditional posterior distribution for each parameter to be estimated. Under general conditions, when the procedure is iteratively executed, the resulting chain of parameters can be considered as a random sample from the joint posterior distribution of the parameters.

For a multilevel model with the adjustment of measurement errors, each iteration of the algorithm consists of eight steps:

Step 1: Sample new set of fixed parameters ($\boldsymbol{\beta}$);

Step 2: Sample new set of level 2 residuals ($\boldsymbol{u}_j$);

Step 3: Sample new level 2 variance ($\boldsymbol{\sigma}_u^2$);

Step 4: Sample new level 1 variance ($\boldsymbol{\sigma}_e^2$);

Step 5: Sample new true values P_l from

$$p(p_{l(ij)}|Y, p_l^*, \beta, \sigma_e^2, u_{0j}, u_{lj}, \sigma_{m_l}^2, \theta, \emptyset^2) \sim N(\hat{p}_{l(ij)}, \hat{V}_{l(ij)})$$

$$\hat{V}_{l(ij)} = \left[(\beta_l + u_{lj})^2 \sigma_e^{-2} + \sigma_m^{-2} + \emptyset^{-2}\right]^{-1}$$

$$\hat{p}_{l(ij)} = \hat{V}_{l(ij)}\left[(\beta_l + u_{lj})\left(y_{ij} - (\beta_0 + u_{oj} + \beta_5 D_{ij})\right)\sigma_e^{-2} + p_{l(ij)}^* \sigma_m^{-2} + \theta\emptyset^{-2}\right]$$

If l=3, after sampling the true value of $p_{3(ij)}$, compute the true quadratic term, $p_{3(ij)}^2$, to include it in the data matrix;

Step 6: Sample a new mean for the true values from

$$p(\theta_l|p_{lij}, \emptyset^2) \sim N(\hat{\theta}_l, \hat{V}_\theta)$$

$$\hat{\theta}_l = \mathrm{n}^{-1}\sum_{\mathrm{ij}} \mathrm{p}_{\mathrm{l(ij)}}$$

$$\hat{V}_\theta = \emptyset^2 \mathrm{n}^{-1}$$

where n is the number of level 1 units;

Step 7: Sample new variance parameter for the true values ($\emptyset^2$);

Step 8: Compute the level 1 residuals ($\boldsymbol{e}_{ij} = y_{ij} - (X_{ij}\hat{\beta} + \hat{u}_j)$).

The estimates presented here consider a burn-in of 500 iterations and chain with 3000 iterations.

Results and Discussion

Two analyses of the model for primary and lower secondary education data were conducted, showing the effects of considering the adjustment for measurement error in the predictor variables of prior achievement. In analysis A the reliability is considered to be R=1 and this produces the unadjusted results. In the analysis B the reliability is considered to be R=0.8 for each prior grade score *t*, *t-1* and *t-2*.

Table 1. Unadjusted Estimates

	Primary Education Estimate (Standard error)	
Fixed Parameters	A: R=1	B: R=0.8
Intercept	-0.041 (0.097)	-0.054 (0.101)
Grade 1 score	0.120 (0.059)	0.115 (0.083)
Grade 2 score	0.223 (0.067)	0.245 (0.106)
Grade 3 score	0.351 (0.057)	0.409 (0.084)
Grade 3 score ^2	0.071 (0.032)	0.092 (0.047)
SES	0.177 (0.060)	0.182 (0.061)
Random Parameters		
Level 1	0.506 (0.053)	0.471 (0.057)
Level 2	0.132 (0.058)	0.138 (0.075)

Table 2. Adjusted Estimates

	Lower Secondary Education Estimate (Standard error)	
Fixed Parameters	A: R=1	B: R=0.8
Intercept	-0.086 (0.107)	-0.124 (0.111)
Grade 7 score	0.417 (0.094)	0.572 (0.194)
Grade 8 score	0.166 (0.085)	0.078 (0.171)
Grade 9 score	0.285 (0.084)	0.333 (0.113)
Grade 9 score ^2	0.000 (0.038)	0.021 (0.049)
SES	0.069 (0.056)	0.067 (0.058)
Random Parameters		
Level 1	0.572 (0.075)	0.536 (0.087)
Level 2	0.065 (0.050)	0.0734 (0.061)

Table 1 summarizes the estimates obtained from primary education data. At this cycle of studies all the prior grade scores included are positive and statistically significant. It can be observed that the magnitude of the estimate tends to decrease as the time lag increases. Concerning the impact of the measurement error adjustment, with exception of estimate $\hat{\beta}_1$, all other estimates have their values increased when the adjustment for measurement error is made, as expected. Thus, for two students in the same class and with the same SES, the effect per point of difference in their grade one scores is, on average, 0.115 points on the outcome score at grade 3. The effect per point of difference in their grade 2 and grade 3 scores on the outcome is 0.245 and 0.409, respectively. The coefficient of the quadratic term is also statistically significant. The relationship between pupil's socioeconomic status and his/her performance at the end of grade 3 is statistically significant ($\hat{\beta}_5$=0.182).

The estimates presented in table 2, for lower secondary education model parameters, from the substantive point of view show a pattern quite different from that one of primary education. However, from the statistical point of view, the comments above mentioned remain valid. That is, in general, estimates of fixed parameters tend to increase as the measurement error is considered. The magnitude of the parameter estimate related to prior grade 7 score must be noted. After adjustment, its effect on the outcome at grade 9 is $\hat{\beta}_1 = 0.572$, on average, controlled by all other variables. It suggests that the pupil's prior achievement at the entrance to lower secondary education works like a "filter" to his/her success at the end of this cycle of education. In addition, the model suggests that the parameter representing the relationship between student's socioeconomic status and the achievement at the end of the cycle (which is the end of compulsory education) is not statistically significant.

Level 1 variance estimates are reduced with the adjustment. Level two variance estimates change relatively little as a result of adjustment compared with those of level 1. Every standard error estimates increase when the adjustment is made.

Since mathematics tests administered to students enrolled in contiguous grades included repeated items (for the purpose of scale equating), the assumption of independent measurement errors should be investigated. Specifically, a lag one serial correlation should be checked. Further analyses are in progress.

REFERENCES

Ballou, D., Sanders, W., Wright, P. (2004). Controlling for Student Background in Value-Added Assessment of Teachers. *J. Educ.Behav. Stat.* 29, 37-65.

Braun, H., Wainer, H. (2007). Value-Added Modeling. In: Rao, C.R., Sinharay, S. (eds.) *Handbook of Statistics* 26, Psychometrics, pp. 475–501. Elsevier.

Browne, W., Goldstein, H., Woodhouse, G. and Uang, M. (2001). An MCMC algorithm for adjusting for erros in variables in random slopes multilevel models. *Multilevel modeling newsletter,* 13, 4-9.

Carrol R.J., Ruppert D., Stefanski L.A., Crainiceanu, C.M. (2006). *Measurement error in nonlinear models,* 2nd edition, Chapman & Hall / CRC.

Ferrão, M.E., Costa, P., Dias, V., Dias, M.(2006). Medição da competência dos alunos do ensino básico em Matemática: 3EMat, uma proposta. [Measuring math skills of students in compulsory education: 3EMat, a proposal]. In: *Actas da XI Conferência Internacional de Avaliação Psicológica* [Proceedings of the XI International Conference on Psychological Evaluation] Braga, Portugal.

Ferrão, M.E., Goldstein, H. 2009. Adjusting for Measurement Error in the Value Added Model: evidence from Portugal. *Quality and Quantity* 43: 951-963.

Fielding, A., Yang, M., Goldstein, H. (2003). Multilevel ordinal models for examination grades. *Stat. Model.* 3,127–153.

Fuller W.A. (2006). *Measurement Error Models*. Wiley: New York.

Goldstein, H.(1997). Methods in school effectiveness research. *School Effectiv. School Improv*. 8, 369–395.

Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edition. Edward Arnold, London.

Goldstein, H., Spiegelhalter, D.J.(1996) League tables and their limitations: statistical issues in comparison of institutional performance. *J. Roy. Stat. Soc*. A 159, 385–443.

Goldstein H, Kounali A and Robinson A. (2008). Modelling measurement errors and category misclassifications in multilevel models. *Stat. Model.* 8, 243-261. DOI: 10.1177/1471082X0800800302.

Ladd, H.F., Walsh, R.P. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review,* 21, 1-17.

Leckie, G. and Goldstein, H. 2009. The limitations of using school league tables to inform school choice. *J. Roy. Stat.* Soc. A, 172: 835-851.

Raudenbush, S.W. and Willms, J.D. (1995). The estimation of school effects. *J. Educ.Behav. Stat.* 20, 307–335.

Vicente, P. (2007). Plano amostral do projecto 3EM—Eficáca Escolar no Ensino da Matemática In: Ferrão, M.E., Nunes, C., Braumann, C. (eds.) *Estatística: Ciência Interdisciplinar,* Proceedings of the XIV Annual Conference of the Portuguese Statistical Society, pp. 849-858. SPE, Lisboa, Portugal.

Woodhouse G., Yang M, Goldstein H and Rasbash J (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society* A, 159, 201-212.