Combining Modelling Strategies to Analyse Teaching Styles Data

NEIL H. SPENCER

Department of Statistics, Accounting and Management Systems, University of Hertfordshire, Hertford Campus, Mangrove Road, Hertford, SG13 8QF, UK. Telephone: +44 1707 285529, Fax: +44 1707 285489, e-mail: N.H.Spencer@herts.ac.uk

Abstract. This paper combines two estimation procedures: Iterative Generalized Least Squares as used in the software MLwiN; Gibbs Sampling as employed in the software BUGS to produce a modelling strategy that respects the hierarchical nature of the Teaching Styles data and also allows for the endogeneity problems encountered when examining pupil progress.

Key words: endogeneity; iterative generalized least squares; gibbs sampling.

1. Introduction

The progress that pupils make in their academic achievement is seen as an indicator of the quality of the education that they receive. With the possibility of sanctions being imposed on schools and/or teachers that are seen to produce inadequate progress, and role models being made of those schools and/or teachers that produce progress above what is generally expected, it is of prime importance that the assessment of pupil progress is handled with care. This care must be taken in the actual measurement of pupil achievement and also the statistical analyses to which measures are subjected.

The role of multilevel models in the analysis of educational data is now generally accepted. Recently published papers in the field that use them in their analyses include Dryler (1999), Brutsaert (1999), Goldstein & Sammons (1997), Hofman et al (1999). Those that use multilevel analyses to examine pupil progress include Sammons & Smees (1998), Strand (1997), Tymms et al (1997). In the setting of education, pupils are grouped into classes which are then grouped into schools. In the multilevel model, each level of the hierarchy in the data has a random effect associated with it to allow for the contribution to the response of unmeasured or unmeasureable influential factors that operate at each level. Coefficients of regressors in the model may also have random components to allow the effect of the regressors to vary between the different groupings at the various levels of the hierarchy. It is this respect that multilevel models give to the hierarchy of the data that have made their use commonplace. Software specifically designed to fit multilevel models have been available since the 1980s and now include MLwiN (Goldstein et al, 1998), which uses an Iterative Generalized Least Squares algorithm to carry out the estimation; HLM (Bryk et al, 1996) which uses an EM algorithm; VARCL (Longford, 1988) which uses Fisher scoring. Major statistical packages such as SAS (PROC MIXED (SAS Institute, 1992, 1996) which uses a Newton-Raphson algorithm) and S-Plus (nlme function (Mathsoft, 1997) which uses an EM algorithm) also have routines that will fit multilevel models. Standard texts on the subject include Goldstein (1995), Longford (1993), Bryk & Raudenbush (1992), Snijders & Bosker (1999). Good introductory chapters/articles on the subject are Paterson (1991) and Paterson & Goldstein (1991).

The problem of endogeneity is one that is unavoidable in analyses of pupil progress. A typical analysis is one that attempts to relate a student's current achievement (e.g. a current test score) to previous achievement (e.g. a baseline score) (see, for example, Sammons & Smee, 1998). There are several factors that contribute to both the current and earlier test score: ability of pupil, influence of teacher (if the teacher is the same at the time of both tests), influence of school (if the school is the same at the time of both tests). There may also be influences related to the neighbourhood where the pupil lives, the background of the pupil, etc. In the multilevel model, we would expect to see random components of the model that relate to the levels of the data hierarchy: pupil, teacher, class, school. We may also see random components relating to neighbourhood, background, etc. Thus the earlier test score is related to the random components of the multilevel model and the standard regression assumption of independence of the explanatory variables and the random part of the model is not tenable. This fact is largely ignored in the modelling of pupil progress. Papers such as Goldstein & Thomas (1996), Gray et al (1995) and Sammons & Smee (1998) do not address the problems associated with endogeneity with the result that the estimates of their model parameters may not be consistent. This lack of consistency may be of more importance in some situations than in others, but unless the problem is examined, the extent to which the parameter estimates are affected is unknown. Conclusions drawn from the suspect results may be inappropriate.

In this paper, a re-analysis of part of the Teaching Styles data (Bennett, 1976) is conducted. The Teaching Styles project had two aims. One was to see if different types of pupil performed better under certain teaching styles and the other, of prime interest in this paper, was to examine whether the progress of pupils was affected by different teaching styles.

Bennett sent questionnaires regarding classroom techniques to 1500 primary school teachers in 871 schools in Lancashire and Cumbria and used principal components analysis followed by cluster analysis to identify different teaching styles. A seven cluster model was adopted and these seven were subsequently reduced to three clusters. The clusters were identified as containing teachers using Formal, Mixed and Informal teaching methods. To examine pupil progress in relation to the three teaching styles, Bennett chose 12 year four teachers of each of the three styles. The teachers chosen were those that displayed teaching characteristics typical of the styles. Attainment and personality tests were then administered to the pupils in these teachers' classes in the September at the start of their fourth year and again at the end of the academic year the following June.

The analysis of pupil progress in three subject areas (Reading, Mathematics and English) was presented in Chapter 5 of Bennett (1976) using analysis of covariance. The use of this method of analysis ignores the clustering of the pupils into classes and effectively assumes that each pupil was taught by a different teacher, leading to small standard errors. Aitkin, Anderson & Hinde (1981) presented a re-analysis of the data (see also Aitkin, Bennett & Hesketh, 1981; Gray & Satterly, 1981), and for pupil progress used a variance components modelling approach that respected the hierarchical nature of the data (pupils grouped under teachers) and also defined the styles of the teachers using a latent class model.

Spencer & Fielding (1998) carried out a comparison of modelling strategies for value-added analyses of educational data using the Iterative Generalized Least Squares algorithm found in the package MLn (Rasbash & Woodhouse, 1995) and a Bayesian approach using Gibbs sampling as found in the BUGS package (Spiegelhalter et al, 1995). In this current work, aspects of both these approaches are used together to analyse the Teaching Styles data, with the package MLwiN (Goldstein et al, 1998) used as the successor to MLn. The work in this paper concentrates on obtaining an easily implemented estimation strategy that employs readily available modelling software so that the methods are capable of being routinely used by researchers carrying out similar analyses.

In section 2 of this paper, models used by Bennett (1976) and Aitkin, Anderson & Hinde (1981) are introduced and the problem of endogeneity is recognised. In section 3, two estimation methods that allow for the endogeneity problem (using instrumental variables and using Gibbs sampling techniques) are discussed. An approach is then developed that combines the estimation strategies available via the BUGS and MLwiN software packages. In section 4, the results obtained when the combined approach is used are considered and the specification of prior distributions is discussed in section 5. Conclusions are provided in section 6.

2. Models and the problem of endogeneity

The simple model used by Bennett (1976) relates the post-test score (June test) to the three teaching styles used, allowing for the pre-test score (September test). The model used is thus of the form below.

$$y_{ijk} = \beta_0 + \alpha_j + \beta_1\left(x_{ijk} - \bar{x}\right) + e_{ijk}$$

where y_{ijk} is the post-test score for pupil k taught by teacher i using teaching style j and x_{ijk} is the pre-test score for the same pupil. The α_j is the fixed effect of teaching style j and e_{ijk} is the pupil-specific random error. This model effectively combines any and all the effects the teachers have into just the teaching style effect, α_j, and thus assumes that all teachers of a particular style are identical. Also because no allowance is made for the clustering of pupils within classes under a single teacher, the model effectively makes the assumption that each pupil has his or her own individual teacher, not shared with any of the other pupils. These assumptions that are made by adopting this model are clearly not tenable.

Aitkin, Anderson & Hinde (1981) introduced a variance component for the teacher into the model, as below (with a simple change in notation).

$$y_{ijk} = \beta_0 + \alpha_j + \beta_1\left(x_{ijk} - \bar{x}\right) + \delta_{0i} + e_{ijk}$$

where δ_{0i} is the random effect of teacher i. This introduction of this teacher effect allows teachers to vary in the effects they have on their pupils' scores beyond simply the effect of teaching style and also enables pupils within the same class to have the same teacher effect applied to them, thus acknowledging the hierarchical structure of the data. This model is therefore a considerable improvement over that used by Bennett (1976).

In this paper we now additionally recognise that the contribution of the pupil to the post-test score is e_{ijk} (subsumed with the lowest level error component) and that the regressor, x_{ijk}, being a pre-test score on the same pupil, will contain a contribution from the pupil that will have a positive correlation with e_{ijk}. We also additionally recognise that the teacher effect for the post-test, δ_{0i}, also accounts for any school effects that influence the post-test. As these school effects may also influence the pre-test score, the regressor x_{ijk} will also contain a contain a contribution from the school that will have a positive correlation with δ_{0i}. Thus x_{ijk} is an endogenous variable for the model, being correlated with both random components, δ_{0i} and e_{ijk}, and we must consider the effect that this has on the estimation procedure used.

Standard multilevel estimation procedures, such as those used in the software described in section 1, make the assumption that the regressors in a model are independent of the random effects and are thus not designed to produce consistent estimates when applied to models with an endogenous variable. Usually this will mean that the estimates obtained for a model with an endogenous regressor will be inconsistent. There is, however, a set of circumstances when this is not the case. If the endogeneity is only caused by correlations at levels of the model hierarchy above the lowest level and the assumption of normality is made for all components of the random part of the model, then it can be shown that, subject to a reparameterisation, the conditional distributions allowing for the endogeneity are equivalent to a specification of the model using conditional distributions assuming that the endogeneity is not present (Aitkin, 1999). In these circumstances, the standard estimation procedures would not produce inconsistent results.

For the modelling of the Teaching Styles data in this paper, these set of circumstances would exist if the endogeneity was only due to the correlation of the regressor, x_{ijk}, with the teacher effect, δ_{0i}. However, as the endogeneity is also caused by the correlation with the pupil effect at the time of post-test, e_{ijk}, we have an added complication and inconsistent parameter estimates may result. The extent to which this lack of consistency affects the parameter estimates will depend greatly on the sizes of the correlations involving the regressor which cause the endogeneity. As, however, these correlations are not known and can only be speculated about, failure to address the endogeneity problem will have an unknown impact on the estimates and the conclusions subsequently drawn. Clearly some approach to overcome the endogeneity problem is required.

3. Modelling Strategies

3.1. USING MLWIN WITH INSTRUMENTAL VARIABLE ESTIMATION

The above model suggested by Aitkin, Anderson & Hinde (1981) can easily be used to model the Teaching Styles data and parameter estimates can be obtained using a multilevel modelling package such as MLwiN.

However, on their own, the techniques used by MLwiN and other packages similarly designed for multilevel modelling do not address the problem of endogeneity. A possible solution is offered by using instrumental variable methods which were first developed to address precisely this problem of correlation between the regressors and model disturbance. In essence, an instrumental variable estimation procedure uses a set of variables that act as "instruments" for the original set of regressors. This instrument set may include some of the original regressors but the key is that the instrument set should be uncorrelated with the model disturbance while at the same time being closely related to the original regressor set. The instrument set is then used in the estimation process alongside the original regressor set in such a manner that consistent estimates result.

Originating and widely used in the field of econometrics, these instrumental variable methods have been the subject of some scepticism. The major concern is that of defining an appropriate instrument set. If the original regressor set is correlated with the model disturbance then is likely that any other set of variables closely related to it (as the instrument set is required to be) will also be correlated with the model disturbance. It is this choosing of the variables to act as the instrument set that is crucial to the success and integrity of the instrumental variable approach. This, in turn, leads to another problem associated with instrumental variable methods. If an instrument set is chosen so that its lack of correlation with the model disturbance is beyond question, then it may not be as closely related to the original regressor set as one would wish. This lack of a close correlation between instrument and regressor sets can lead to large standard errors for the instrumental variable estimates and, on some occasions, leave the estimates uninterpretable. For further information on instrumental variable methods see, for example, Bowden & Turkington (1984).

Previous work concerning the application of instrumental variable methods to multilevel models with endogenous variables has been carried out by Fielding & Spencer (1997, 1998). In their work, consistent estimates of the fixed parameters in the model are first obtained using instrumental variables. An appropriate multilevel modelling package such as MLwiN is then used to estimate the other model parameters (the random parts of the model) whilst constraining the fixed parameters to be equal to the instrumental variable estimates. The resulting estimates of the random parameters of the model are then used to obtain the standard errors of the instrumental variable estimates.

The reason why this strategy would produce consistent parameter estimates can be seen by examining why inconsistent estimates are arrived at if no steps are taken to address the endogeneity problem. The Iterative Generalized Least Squares algorithm in MLwiN is, as its name suggests, an iterative procedure with firstly estimates of the fixed part of the model being obtained and then these are used to obtain estimates of the random part of the model. These estimates of the random part are then used to re-estimate the fixed part of the model and the iterative procedure continues until convergence of the parameter estimates occurs. In a situation where endogeneity exists, the initial estimate of the coefficient of the endogenous variable will not be consistent. This inconsistent estimate will then be used in the estimation of the random part of the model and cause further inconsistency here. This inconsistency of the estimates of the random part of the model will then be fed back into the estimation of all the parameters in the fixed part of the model. By using the consistent instrumental variable estimates of the fixed effects in the algorithm, the random parameters will be estimated consistently and so all the parameter estimates that result will be consistent.

In order to implement this strategy in the context of the model used in this paper, an instrument must be found that is related to the endogenous pre-test score whilst being independent of the model disturbance. This is not always straightforward as it is often the case that such suitable background variables have not been collected. With the Teaching Styles data, there is a wealth of information regarding personality attributes available, measured at both the pre-test and post-test occasions. The post-test personality measurements may clearly be related to the post-test model disturbance (which includes a pupil-specific effect), and are thus unsuitable for use as instruments. The relationships between the pre- test personality measurements and the post-test model disturbance is less clear and need further investigation before they can be considered suitable for use as instruments. This is the subject of ongoing work and for the purposes of this paper, the cautious option of not using them as instruments is chosen. As a result, the Teaching Styles data does not have available obvious variables that are appropriate for use as instruments with which to implement the instrumental variable approach, so alternatives must be considered.

3.2. USING BUGS

An alternative modelling strategy is to use a completely Bayesian approach, as implemented via Gibbs sampling in the package BUGS. A model can be constructed which respects the hierarchical nature of the data, as below.

$$y_{ijk} = \alpha_1 z_{1ijk} + \alpha_2 z_{2ijk} + \alpha_3 z_{3ijk} + \beta_1\left(x_{ijk} - \bar{x}\right) + \delta_{0i} + e_{0ijk}$$

where y_{ijk} is the post-test score for pupil k taught by teacher i using teaching style j and x_{ijk} is the pre-test score for the same pupil. The α_1, α_2, α_3 are the main effects for the three teaching styles with the z_{1ijk}, z_{2ijk}, z_{3ijk} being appropriately defined dummy variables. The δ_{0i} is the random effect associated with teacher i and the e_{0ijk} is the pupil-specific effect. This model is identical (subject to a change in parameterisation) to that used by Aitkin, Anderson & Hinde (1981).

The endogeneity of the pre-test score can also be respected by allowing it to be modelled as below.

$$x_{ijk} - \bar{x} = \delta_{1i} + e_{1ijk}$$

where δ_{1i} is the random effect associated with teacher i and the e_{1ijk} is the pupil-specific effect.

To respect the endogeneity, the δ_{0i}, δ_{1i} are defined as coming from a bivariate normal distribution and, independently, the e_{0ijk}, e_{1ijk} are also defined as coming from a bivariate normal distribution.

Non-informative priors (normal distributions with means of zero and precisions of 0.00001) are assigned to the fixed parameters. These effectively say that before examination of the data, the effects of the teaching styles are assumed to be identical and zero, but the low precisions and thus high variances attached to this assumption mean that it is a very weak assumption and information from the data will swamp its effect. For the bivariate normal distributions for the random parameters, we have zero mean vectors as priors (i.e. random effects have means of zero) and arbitrarily defined scale matrices with the numbers 2 on the main diagonal and 1 in the off-diagonal positions. These matrices indicate the relative sizes of the covariance matrices. Other definitions of the prior distributions could be used and are investigated later in section 5 of this paper.

A BUGS analysis starts with each parameter having a pre-determined start value. A Gibbs sampling algorithm then simulates the full joint distribution of all the parameters by obtaining a single sample for each parameter, one at a time, from its conditional distribution given all the other parameters (see Best, Spiegelhalter, Thomas & Brayne, 1996, for more details). To begin with, the starting values will have a non-negligible influence on the sample values obtained for each parameter. However, as the sampling continues, the effect of these starting values will diminish and eventually effectively disappear. At some time after the commencement of the Gibbs sampler, convergence should occur. This effectively means that the conditional distributions, from which samples are being obtained, are not changing from iteration to iteration of the sampler. Once this convergence has occurred, the subsequent samples obtained from the Gibbs sampler can be seen as being realisations of the joint distribution and sample statistics (e.g. mean, standard deviation) can be calculated from them to summarise the distributions of the parameters.

There remains the question of how to decide when then parameter values have converged. Graphical displays of the updated parameter values can be inspected and there are also convergence criteria that have been developed. Accompanying the BUGS software is a suite of menu-driven S-Plus functions known collectively as CODA (Best, Cowles & Vines, 1995) which helps the BUGS user in determining whether or not convergence has occurred.

With the help of CODA, a run of 500 BUGS updates was considered for the Reading tests from the Teaching Styles data. Although being a relatively large number of updates, taking some considerable computer time, inspection of plots of the updated parameter values showed that they did not all display the pattern of white noise around a stationary value, characteristic of a state of convergence. Use of the convergence criteria available through CODA also revealed a lack of convergence for some of the parameters.

Although convergence for all the parameter estimates was not occurring, that of β_1, the coefficient of the endogenous variable, did appear to have converged after a reasonably short period of time. This was also true when the Mathematics and English tests from the Teaching Styles data were examined. This prompted the consideration of a combined strategy for modelling the data.

3.3. A COMBINED STRATEGY

This combined strategy uses the multilevel modelling abilities of the package MLwiN, with its flexibility for constraining parameters, with the ability of BUGS to obtain a consistent estimate of the coefficient of the endogenous variable from an analysis that takes a modest amount of time to run.

In a similar manner to the way the instrumental variable estimates of the fixed effects could be used as described in section 3.1, MLwiN is used to estimate the parameters of the multilevel model whilst constraining the coefficient of the endogenous regressor to be equal to the value obtained from the BUGS analysis. The initial estimate of the coefficient of the endogenous variable will be consistent and so via the Iterative Generalized Least Squares algorithm the other fixed effects will also have consistent estimates. This will then yield consistent estimates of the random part of the model and the iterative procedure will continue until convergence. What then result are consistent estimates of all the model parameters: that for the coefficient of the endogenous variable coming from the BUGS analysis and those for the other parameters coming from the MLwiN analysis. Standard errors produced by MLwiN will be conservative as they do not allow for variation in the coefficient of the endogenous variable and this must be borne in mind when conclusions are drawn.

We thus have a means of obtaining consistent parameter estimates for all parts of the model using a combination of the modelling strategies used with BUGS and MLwiN.

4. Comparison of results

4.1. USING TEACHING STYLES AS DEFINED BY BENNETT (1976)

The combined modelling strategy was employed in analysing the Reading, Mathematics and English tests from the Teaching Styles data, with the allocation of teachers to styles being that defined in Bennett (1976). The length of the BUGS run used was 250 updates. Convergence of β_1 appeared to have been achieved after 100 updates and its mean value was calculated from the last 150 updates. This was a cautiously large number of updates and a shorter run may also have produced acceptable results.

We present the results obtained using the combined BUGS and MLwiN strategy for estimates of α_1, α_2 and α_3, the main effects for the teaching styles, in table I (with standard deviations in brackets). For comparison purposes, we also have figure 1 (following the format of figures from Aitkin, Anderson & Hinde, 1981) showing these results from the combined strategy alongside the effects of teaching styles published in Bennett (1976). Also displayed are estimates of the effects of teaching styles obtained from an EM algorithm similar to that used by Aitkin, Anderson & Hinde (1981) whose paper does not provide such information for the teaching styles as defined by Bennett (1976). Notably, we have standard errors to consider here that were not provided by Bennett (1976), and the EM algorithm used by Aitkin, Anderson & Hinde did not provide them either.

Table I. Main effects for Bennett's teaching styles from combined strategy

	Formal style	Mixed style	Informal style
Reading	105.3 (1.387)	106.1 (1.397)	104.7 (1.341)
Mathematics	103.8 (1.191)	102.7 (1.210)	103.2 (1.160)
English	107.7 (0.936)	106.6 (0.949)	105.6 (0.910)

As can be seen from figure 1, for English, the three estimation strategies produce very similar results. For Mathematics, the EM algorithm and Bennett's estimates are quite close but the combined strategy produces smaller differences between the teaching styles, although the ordering of the estimates is maintained. For Reading, the combined strategy produces estimates that have the same order as Bennett's estimates but the differences are less extreme. The EM algorithm produces a different picture with the ordering of the Formal and Mixed styles reversed.

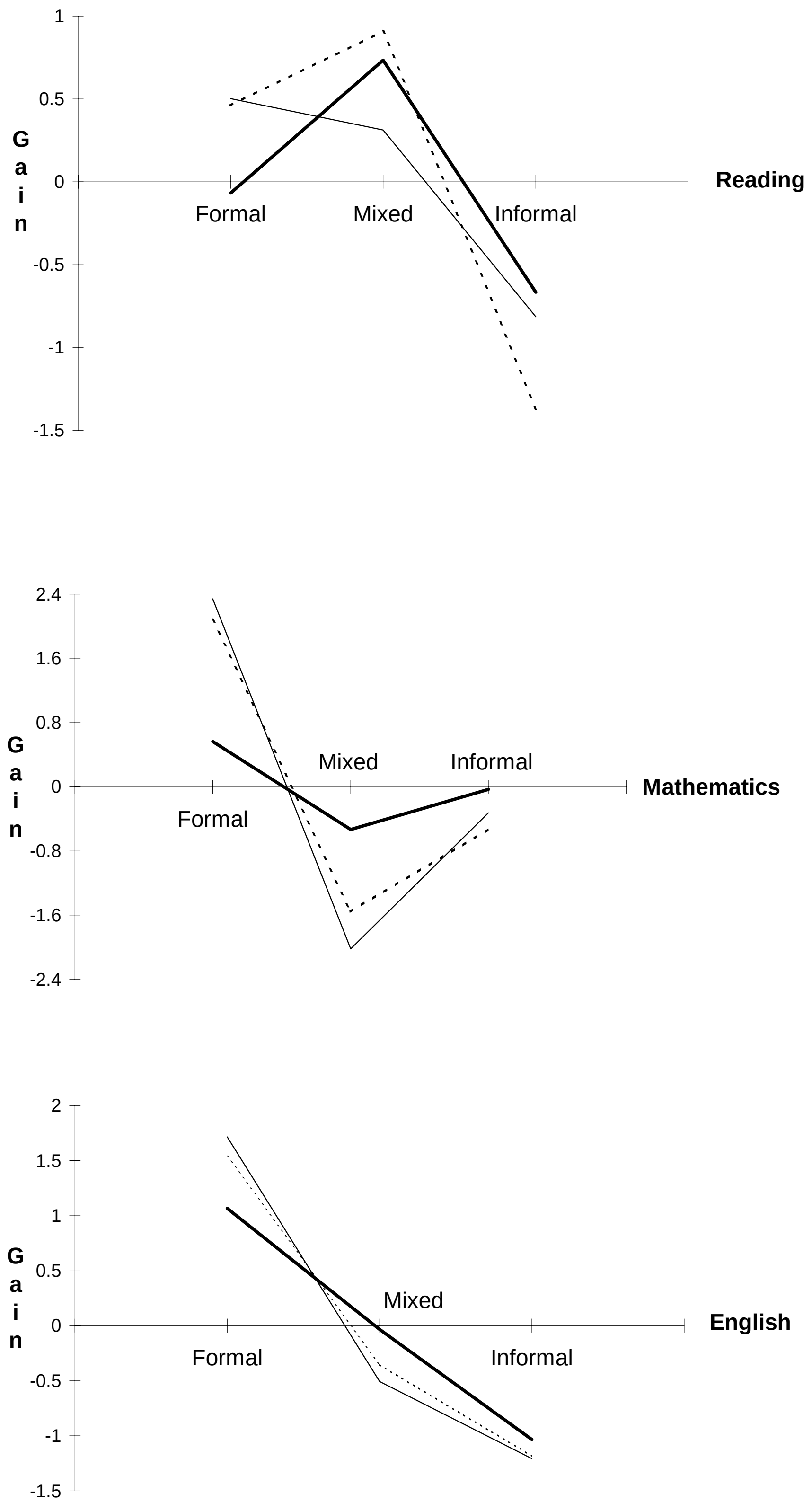

Figure 1. Teaching style effects for combined strategy, Bennett (1976) and EM algorithm

- - - Bennett (1976), —— EM algorithm, ━━ Combined Strategy

Looking at table I, we see that for Reading and Mathematics, the differences between the estimates for the styles is within two standard deviations leading us to observe that there is insufficient evidence to claim that any style is better or worse than the others regarding the results of the post-test, using the two standard deviations rule of thumb. For English it appears that the estimates for the Formal and Informal styles are more than two standard deviations apart. However, it should be noted that the standard deviations given are conservative ones at best because when the MLwiN part of the combined strategy was being undertaken, the variation associated with the BUGS estimate of the β_1 coefficient was not used. Thus the real standard errors for the style effects should be greater than those displayed in table I and the difference between the Formal and Informal styles for English may not be as important as it first seems.

4.2. USING TEACHING STYLES AS DEFINED BY AITKIN, ANDERSON & HINDE (1981)

The combined modelling strategy was also used to analyse the Reading, Mathematics and English tests from the Teaching Styles data, with the allocation of teachers to styles now being that defined by Aitkin, Anderson & Hinde (1981). The estimate of the coefficient of the endogenous variable from BUGS was obtained in the same manner as in section 4.1.

Table II shows the main effects for the teaching styles obtained using the combined BUGS and MLwiN strategy (with standard deviations in brackets). We also have figure 2 showing these results from the combined strategy alongside those from Aitkin, Anderson & Hinde (1981).

Table II. Main effects for Aitkin, Anderson and Hinde's teaching styles from combined strategy

	Formal style	Mixed style	Informal style
Reading	104.8 (1.172)	104.2 (1.816)	106.2 (1.394)
Mathematics	103.3 (0.976)	101.5 (1.503)	103.8 (1.162)
English	107.4 (0.794)	104.8 (1.225)	106.1 (0.946)

As can be seen from figure 2, the ordering of the estimates for the three teaching styles is the same for Aitkin, Anderson & Hinde (1981) and the combined strategy for all three subjects. Importantly though, the combined strategy produces estimates that show less extreme differences between the styles. With the same warning about the standard errors shown in table II as described in section 4.1, we observe that there is insufficient evidence to claim that the teaching styles differ in their effects using the two standard deviations rule of thumb.

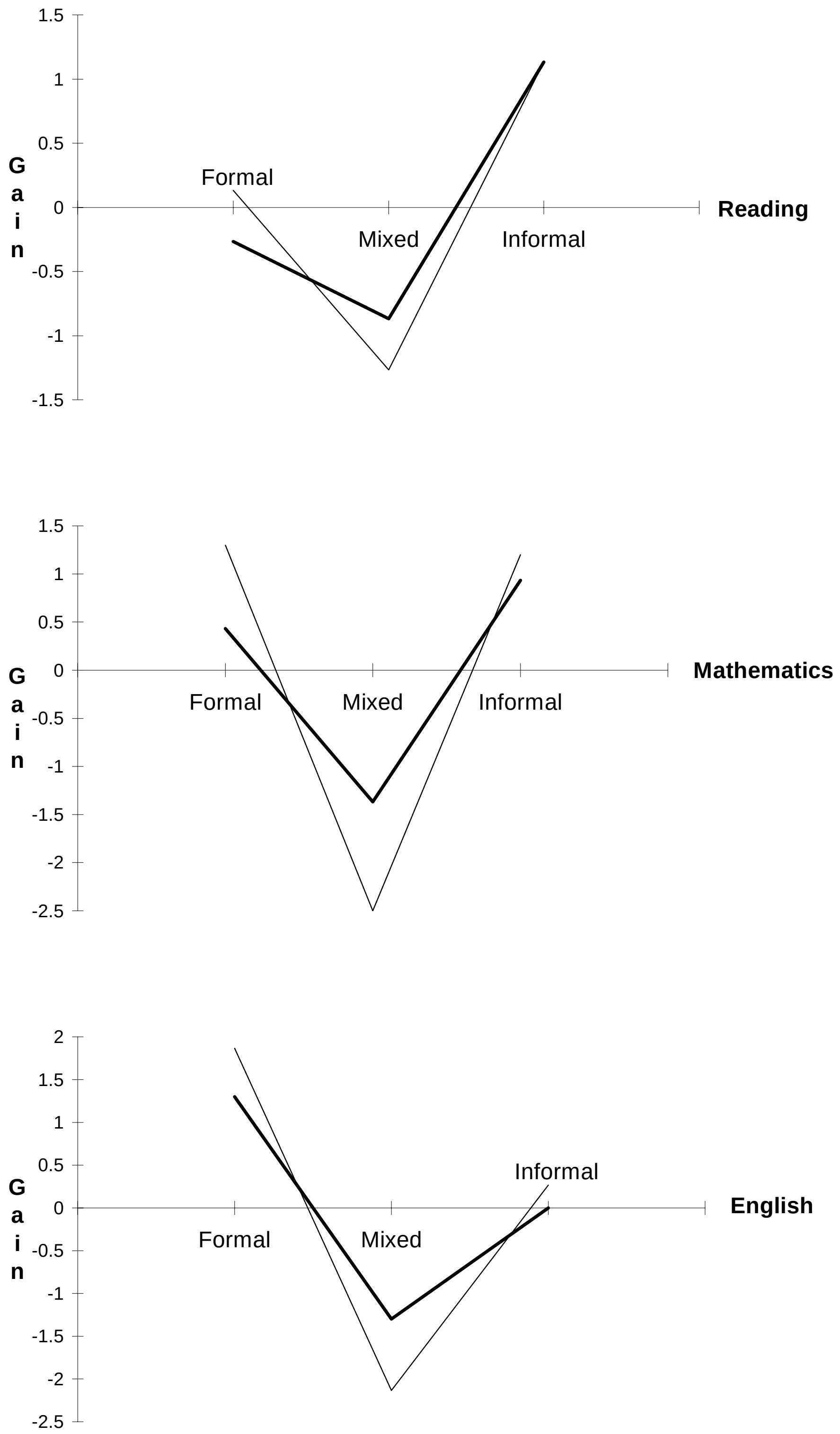

Figure 2. Teaching style effects for combined strategy and Aitkin, Anderson & Hinde (1976)

Aitkin, Anderson & Hinde (1981), Combined Strategy

Despite the fact that the results from the combined strategy do not differ greatly from those obtained by Bennett (1976), they have been arrived at by properly taking into account the hierarchical nature of the data and also the problem of endogeneity. Therefore, it is possible to interpret the results with a greater degree of confidence, and given the sizes of the standard deviations, the conclusion is reached that there appear not to be any significant differences between the main effects of the three teaching styles.

5. Priors

In the BUGS analyses carried out above, non-informative priors were used. To briefly assess the sensitivity of the estimate of the coefficient of the pre-test score to different specifications of the prior distributions, two further BUGS runs were undertaken using the Reading test data. The first run simplified the approach to the scale matrices for the bivariate normal distributions for the random parameters by using identity matrices. The second run used different but sensible prior distributions for the α_1, α_2, α_3 (normal distributions with means equal to the mean of the post-test and precisions of 0.00001) and for β_1 (a normal distribution with mean equal to 1 and precision of 0.00001).

As before, runs of 250 updates were conducted and convergence of β_1 appeared to have been achieved after 100 updates. Its mean value was calculated from the last 150 updates. Both the runs yield a mean value of 0.837 for β_1, close to the 0.874 used previously. Thus, although a comprehensive review of alternative prior distributions has not been carried out, some degree of confidence can be felt in the robustness of the estimate to alternative specifications of the prior distributions.

6. Conclusions

The analysis of the Teaching Styles data shown in this paper has revealed that there is not enough evidence to suggest that real differences exist between the results obtained from the three styles of teaching, whether the teachers are allocated to styles by the method of Bennett (1976) or that of Aitkin, Anderson & Hinde (1981). This finding, coming from an analysis that both respects the clustering of the pupils in the study and the existence of an endogenous variable in the modelling process unlike previous analyses, is made possible due to the availability of (admittedly conservative) standard errors for the estimates of the teaching style effects that had not been available for previous analyses.

It has been shown that the modelling strategies employed in BUGS and MLwiN can be combined to provide the researcher with a means of obtaining consistent parameter estimates for models containing an endogenous variable. The extent to which ignoring the endogeneity problem affects the results of an analysis is not known before the problem is addressed and may substantially affect the conclusions drawn. Combining the approaches overcomes difficulties encountered when attempting to carry out the analysis using instrumental variables with MLwiN (lack of suitable variables to be used as instruments) and when using BUGS on its own (problems of convergence). As always, the researcher must be competent at using the packages involved and have a working knowledge of the procedures that underpin them. This is particularly the case for the BUGS side of the combined strategy where prior distributions must be specified and convergence judged. It is also noted that some Gibbs sampling procedures are available through currently available versions of MLwiN and it is possible that in the future the combined strategy may be available using just this package, although the level of understanding required of the researcher will be the same as when BUGS is used.

Acknowledgements

The author would like to thank Murray Aitkin and John Hinde for their comments on a conference paper version of this article given at the 14th International Workshop on Statistical Modelling, Graz, Austria, July 1999, and Tony Fielding for his comments on an earlier version of this paper.

References

Aitkin, M. (1999). Personal communication after conference presentation and referee’s report on a paper in a related area.

Aitkin, M., Anderson, D. & Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society A*, **144**, 4, 419-461.

Aitkin, M., Bennett, S.N. & Hesketh, J. (1981). Teaching styles and pupil progress - a reanalysis. *British Journal of Educational Psychology*, **51**, 170-186.

Bennett, N. (1976). *Teaching Styles and Pupil Progress*. London: Open Books.

Best, N.G., Cowles, M.K. & Vines, S.K. (1995). *CODA Manual, Version 0.30*. Cambridge: MRC Biostatistics Unit.

Best, N.G., Spiegelhalter, D.J., Thomas, A. & Brayne, C.E.G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society, Series A*, **159**, 2, 323-342.

Bowden, R.J. & Turkington, D.A. (1984). *Instrumental Variables*. Cambridge: Cambridge University Press.

Brutsaert, H. (1999). Coeducation and gender identity formation: a comparative analysis of secondary schools in Belgium. *British Journal of Sociology of Education*, **20**, 3, 343-353.

Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.

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

Dryler, H. (1999). The impact of school and classroom characteristics on educational choices by boys and girls: a multilevel analysis. *Acta Sociologica*, **42**, 4, 299-318.

Fielding, A & Spencer, N.H. (1997). Modelling cost-effectiveness in general certificate of education A level: Some practical methodological innovations. In Minder, C.E. & Friedl, H. (eds), *Good Statistical Practice: Proceedings of the 12th International Workshop on Statistical Modelling, Biel/Bienne, July 7 to 11, 1997*. Schriftenreihe der Österreichischen Statistischen Gesellschaft: Vienna.

Fielding, A. & Spencer, N.H. (1998). Value-added analysis for cross-classified multilevel data. *University of Birmingham, Department of Economics Discussion Paper 98-15.*

Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G. & Healy, M. (1998). *A user's guide to MLwiN, Version 1.0.* London: Institute of Education, University of London.

Goldstein, H. & Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis. *School Effectiveness and School Improvement*, **8**, 219-230.

Goldstein, H. & Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society, Series A*, **159**, 1, 407-442.

Gray, J., Jesson, D., Goldstein, H., Hedger, K. & Rasbash, J. (1995). A multi-level analysis of school improvement: changes in schools' performance over time. *School Effectiveness and School Improvement*, **6**, 2, 97-114.

Gray, J. & Satterly, D. (1981). Formal or informal? A re-assessment of the British evidence. *British Journal of Educational Psychology*, **51**, 187-196.

Hofman, R.H., Hofman, W.H.A. & Guldemond, H. (1999) Social and cognitive outcomes: a comparison of contexts of learning. *School Effectiveness and School Improvement*, **10**, 3, 352-366.

Longford, N.T. (1988). *VARCL -- software for variance component analysis of data with hierarchically nested random effects (maximum likelihood)*. Princeton, NJ: Education Testing Service.

Longford, N.T. (1993). *Random Coefficient Models*. New York: Oxford University Press.

Mathsoft (1997). *S-Plus 4: Guide to Statistics*. Seattle: Data Analysis Products Division, Mathsoft, Inc.

Rasbash, J. & Woodhouse, G. (1995). *MLn Command Reference, Version 1.0.* London: Institute of Education, University of London.

Sammons, P. & Smees, R. (1998). Measuring pupil progress at key stage 1: using baseline assessment to investigate value added. *School Leadership and Management*, **18**, 3, 389-407.

SAS Institute (1992). *SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements*. Cary, NC: SAS Institute, Inc.

SAS Institute (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.11*. Cary, NC: SAS Institute, Inc.

Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publishers.

Spencer, N.H. & Fielding, A. (1998). A comparison of modelling strategies for value-added analyses of educational data. In Marx, B. & Friedl, H. (eds.), *Statistical Modeling: Proceedings of the 13th International Workshop on Statistical Modeling, New Orleans, Louisiana, USA, July 27-31, 1998.*

Spiegelhalter, D.J., Thomas, A., Best, N.G. & Gilks, W.R. (1995). *BUGS: Bayesian inference Using Gibbs Sampling, Version 0.50*. Cambridge: MRC Biostatistics Unit.

Strand, S. (1997). Pupil progress during key stage 1: a value added analysis of school effects. *British Educational Research Journal*, **23**, 471-488.

Tymms, P., Merrell, C. & Henderson, B. (1997). The first year at school: a quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*, **3**, 101-118.