Methods of Psychological Research Online 2000, Vol. 5, No. 4
Internet: http://www.mpr-online.de
Institute for Science Education

The Mixed Birnbaum Model: Estimation using Collateral Information

Arnold Smit[1,2]
Henk Kelderman[1]
Henk van der Flier[1]

Abstract

This paper introduces the Mixed Birnbaum model with collateral information. A simulation study is conducted to evaluate parameter estimation for this model. More specifically, the gain of incorporating collateral information into the model is investigated. The results show that the standard errors as well as latent class assignment can benefit substantially from incorporating external variables that associate with the latent class variable, especially when the sample size is relatively large.

Keywords : item response theory; latent class analyses; mixed IRT models

---

[1]Department of Work & Organizational Psychology, Free University Amsterdam; v.d. Boechorststraat 1; 1081 BT, Amsterdam

[2]Netherlands Organization of Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was supported by a grant of the Foundation for Behavioral and Educational Sciences of this organization (575-30-004).

1 Introduction

In the social sciences, sharp definitions of constructs, and therefore clear procedures for classification or quantification, are often difficult to give. In psychology fuzzy concepts like extraversion, need for achievement, or, numeric or verbal intelligence constitute the theoretical building blocks.

To evaluate the validity of psychological theories, we are faced with the problem that for these fuzzy concepts no single unambiguous measure can be constructed. Instead typically a set of J items, with fixed discrete response categories, is constructed. These items should at least elicit some differentiation in responses due to variation in the construct (real valued trait, or discrete classes) measured.

The observed responses to the J items are postulated to be the outcome of some stochastic process. The measurement model is the statistical formalization of the process presumed to underly the response behaviour. If the measurement model adequately summarizes the data, it can be used to estimate the position of subjects on the trait measured. Depending on the assumptions underlying the measurement model, these estimates have certain properties. Stronger assumptions result in measures having stronger measurement properties.

An important class of measurement models consists of the so called item response models, see Rasch (1960), Birnbaum (1968) and Samejima (1969). This class received a lot of attention in the psychometric literature, see for instance Heinen (1993), Fisher and Molenaar (1995), Linden and Hambleton (1997), Wilson, Engelhard Jr., and Draney (1997), and Wilson, Draney, and Engelhard (1999). Item response models can be seen as probabilistic versions of the deterministic model of Guttman (1944).

These probabilistic measurement models provide a framework for inference when using a set of indirect measures. A problem associated with these procedures is the *implicit*, due to the indirect nature of the measures, assumption that the items actually elicit a more or less comparable response process for all subjects. A straightforward procedure to evaluate this assumption is to check if the same model holds for different subgroups (f.i. sexe, race, age). Another, less straightforward, way is to use statistical tools to track unobserved (latent) groups. A particularly interesting framework is that of Latent Class analysis, see Lazarsfeld and Henry (1968), or, more recently Langeheine and Rost (1988), or Hagenaars (1993). In this framework, mostly developed by sociometricians, the goal is to find a partitioning of the sample such that, conditional on the partitioning, simple (independence) structures result that adequately describe the data.

In recent years models with both latent traits and latent classes have emerged, see Rost (1990), Kelderman and Macready (1990), Mislevy and Verhelst (1990), Heinen (1993), and Rost and Langeheine (1997). The Achilles' heel of these models are the large standard errors of the parameter estimates and multi-modality of the likelihood. Smit, Kelderman, and van der Flier (1999) found, with the Mixed Rasch model, that the use of collateral information substantially reduces the standard errors.This paper introduces the Mixed Birnbaum model with a collateral variable associated with the latent class. It investigates the usefullness of incorporating collateral information into the Mixed Birnbaum model, thereby enabling the application of this model to practical situations.

2 The Mixed Birnbaum Model with Collateral Information

Let the random variable $X_{ij} \in \{0,1\}$ denote the score of person i on item j, with realization $x_{ij} = 1$ if the correct response is given, and $x_{ij} = 0$ otherwise. Furthermore, let the random variable $Y_i \in \{0,1\}$ denote an unobserved indicator for the model describing the response behaviour of person i. We assume the sample to be a mixture of two latent groups, whose observed responses are realizations from two distinct Birnbaum models. Finally, let $Z_i \in \{0,1\}$ denote an observed variable that associates with Y_i (collateral or concomitant variable).

The probability of response vector $\boldsymbol{x}_i = \{x_{i1}, \cdots, x_{iJ}\}$ for subject i with ability parameter θ_i, responding according to the Birnbaum model of subpopulation y_i (using the assumption of *local stochastic independence*, is

$$P_{\boldsymbol{x}_i|y_i,\theta_{iy}} \overset{\text{def}}{=} P\left(\boldsymbol{X}_i = \boldsymbol{x}_i|Y_i = y_i, \theta_{iy}, \boldsymbol{\beta}_y, \boldsymbol{\alpha}_y\right) = \prod_{j=1}^{J} \frac{e^{x_{ij}\alpha_{jy}(\theta_{iy}-\beta_{jy})}}{1+e^{\alpha_{jy}(\theta_{iy}-\beta_{jy})}} \tag{1}$$

where $\boldsymbol{\beta}_y = \{\beta_{1y}, \cdots, \beta_{Jy}\}$ and $\boldsymbol{\alpha}_y = \{\alpha_{1y}, \cdots, \alpha_{Jy}\}$ denote the subsample specific item difficulty and item discrimination parameters respectively. It is assumed that $\theta_y \sim N[\mu_y, \sigma_y]$. Now reparameterize using $\theta^* = \frac{\theta_y - \mu_y}{\sigma y}$, then

$$P_{\boldsymbol{x}_i|y_i,\theta_i^*} = \prod_{j=1}^{J} \frac{e^{x_{ij}\alpha_{jy}^*\left(\theta_i^*-\beta_{jy}^*\right)}}{1+e^{\alpha_{jy}^*\left(\theta_i^*-\beta_{jy}^*\right)}}$$

where $\alpha_{jy}^* = \alpha_{jy}/\sigma_y$, $\beta_{jy}^* = \mu_y + \beta_{jy}\sigma_y$, and $\theta^* \sim N[0,1]$ independent of Y. Next the collateral variable Z is incorporated into the model. Using standard probability calculus, the joint probability can be factored as follows

$$P_{\boldsymbol{x}_i,z_i,y_i,\theta_i^*} = P_{\boldsymbol{x}_i|y_i,z_i,\theta_i^*} P_{\theta_i^*|y_i,z_i} P_{y_i,z_i}$$

Assume that all associations between items are explained by the latent trait and the latent class, so $P_{\boldsymbol{x}_i|y_i,z_i,\theta_i^*} = P_{\boldsymbol{x}_i|y_i,\theta_i^*}$. Furthermore, Z and θ^* are associated through Y, or more formally, $P_{\theta^*|y_i,z_i} = P_{\theta^*|y_i}$. Finally, from the reparametrization we have $P_{\theta_i^*|y_i} = P_{\theta_i^*}$. From this it follows

$$P_{\boldsymbol{x}_i,y_i,z_i,\theta_i^*} = P_{\boldsymbol{x}_i|y_i,z_i,\theta_i^*} P_{\theta_i^*|y_i,z_i} P_{y_i,z_i} = P_{\boldsymbol{x}_i|y_i,\theta_i^*} P_{\theta_i^*} P_{y_i,z_i}$$

The model is estimated by maximizing the marginal log-likelihood, where θ^* is integrated out over its assumed standard normal distribution ($P_{\theta_i^*} = \phi(\theta^*)$). The marginal distribution can be approximated using Gauss-Hermite quadrature

$$P_{\boldsymbol{x},y,z} = P_{z,y} \int_{-\infty}^{\infty} P_{\boldsymbol{x}|y,\theta^*}\, \phi(\theta^*)\, d\theta^* \approx P_{y,z} \sum_{k=1}^{K} P_{\boldsymbol{x}|y,T_k}\, A(T_k)$$

where the T_k and $A(T_k)$ denote the abscissas and weights. The likelihood of the sample, assuming Y is known, is multinomial

$$L = \frac{I}{\prod_{\boldsymbol{x}} \prod_y \prod_z e_{\boldsymbol{x}yz}!} \prod_{\boldsymbol{x}} \prod_y \prod_z (P_{\boldsymbol{x},y,z})^{e_{\boldsymbol{x}yz}} \propto \prod_{\boldsymbol{x}} \prod_y \prod_z (P_{\boldsymbol{x},y,z})^{e_{\boldsymbol{x}yz}}$$

where $e_{\boldsymbol{x}yz}$ denotes the frequency of the response vector $\{\boldsymbol{x}, y, z\}$, and I denotes the sample size. For the log-likelihood, say $\ell = \log(L)$, we have

$$\ell \propto \sum_y \sum_z e_{+yz} \log(P_{z,y}) + \sum_y \sum_{\boldsymbol{x}} e_{\boldsymbol{x}y+} \log\left(\sum_{k=1}^{K} P_{\boldsymbol{x}|yT_k} A(T_k)\right) \tag{2}$$

A practical computational procedure for the Birnbaum model, using the EM-algorithm of Dempster, Laird, and Rubin (1977) , was provided by Bock and Aitkin (1981). Apart from the fact that here a logit is used rather than a probit, the likelihood within a latent class is the same. The main difference is the sum over the values of the latent class in the second part of equation 2, and the first part of the likelihood, for which we specify a saturated loglinear model. The model can be estimated by slightly adapting the procedure as proposed by Bock and Aitkin. In the M-step of the algorithm, the first part of the log likelihood is also maximized. Furthermore, in the E-step, the computation of the expected table (as if the latent variables were observed) given the observed data and the current parameter estimates, has an extra component where we expand over the latent class

$$e_{\boldsymbol{x}yz}^{new} = f_{\boldsymbol{x}z} \frac{e_{\boldsymbol{x}yz}^{old}}{e_{\boldsymbol{x}+z}^{old}}$$

where $e_{\boldsymbol{x}yz}^{new}$ and $e_{\boldsymbol{x}yz}^{old}$ denotes the new and old expected frequencies for the complete table respectively, and, $f_{\boldsymbol{x}z}$ denotes the observed frequency. Maximum likelihood estimates can be obtained by repeated application of the E–step and M–step until convergence.

3 Simulation

In order to evaluate the effect of incorporating collateral information into the model, a simulation study is conducted. Data sets with four hundred and two thousand subjects are generated according to the Mixed Birnbaum model, with two equal size classes, using two different parameter sets, as reported in Table 1.

	item									
UNEQUAL	1	2	3	4	5	6	7	8	9	10
β_{j1}	0.0	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
β_{j2}	0.0	-1.0	-1.5	-2.0	-0.5	1.0	0.5	0.0	1.5	3.0
α_{j1}	0.9	0.7	1.1	0.7	1.3	0.9	0.5	1.3	0.5	1.1
α_{j2}	0.9	1.1	0.7	1.1	0.5	0.9	1.3	0.5	1.3	0.7
EQUAL	1	2	3	4	5	6	7	8	9	10
β_{j1}	0.0	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
β_{j2}	0.0	-1.0	-1.5	-2.0	-0.5	1.0	0.5	0.0	1.5	3.0
α_{j1}	0.9	0.7	1.1	0.7	1.3	0.9	0.5	1.3	0.5	1.1
α_{j2}	0.9	0.7	1.1	0.7	1.3	0.9	0.5	1.3	0.5	1.1

Table 1: Parameter values used in simulation

The parameter sets are called EQUAL and UNEQUAL, referring to the equality of the discrimination parameters α_{jy} for the classes. The values of the item difficulty and item discrimination parameters are chosen in the range often encountered with real data. Furthermore, the correlation between the α and β parameters is minimized (0.03). For every sample size $\times$ parameter set combination, one hundred data sets are generated. Thus four hundred data sets are generated. Next, six extra variables are appended to each data set. These collateral variables differ in the degree of overlap with the latent class variable. The bivariate probabilities of the –collateral information $\times$ latent class– cross classification are given in Table 2. The strength of association ranges from independence to equality. The labels of the six collateral variables (50/50,60/40,$\cdots$) refer to the corresponding conditional probabilities.

Association Latent class $\times$ Collateral variable												
	50/50		60/40		70/30		80/20		90/10		equal	
LC 1	.25	.25	.30	.20	.35	.15	.40	.10	.45	.05	.50	.00
LC 2	.25	.25	.20	.30	.15	.35	.10	.40	.05	.45	.00	.50

Table 2: Proportions in Latent class $\times$ collateral variable cross classification

Every data set is analyzed without a collateral variable, in which case we omit the first term from the likelihood of model equation 2, and with each of the appended collateral variables with model. The LEM program, (Vermunt, 1997) , is used to run the analysis. Additionally, data sets generated with the EQUAL parameter set are analyzed with an equality restriction on the discrimination parameters (equal over the latent classes).

4 Results

In the next sections some statistics about the precision of the parameter estimates are reported. These statistics are reported in terms of the difference between the true and estimated parameter values, the deviance. Because of the possibility of converging to local solutions, the fifteen percent highest and lowest deviances are not used. We could, of course, have used the true values as starting values for the analysis, but then we would have lost information about the computational procedure. First, the mean of the deviances is reported. Second, the standard deviations of the deviances are given. Next, some statistics concerning the convergence are given. And finally, the possibility of assigning subjects to one of the latent classes is evaluated.

4.1 Mean of the deviances

The mean of the deviances is computed for the item main and item × latent class interaction parameters aggregated over all items and all replications. First the results for the difficulty parameters are given in Figure 1.

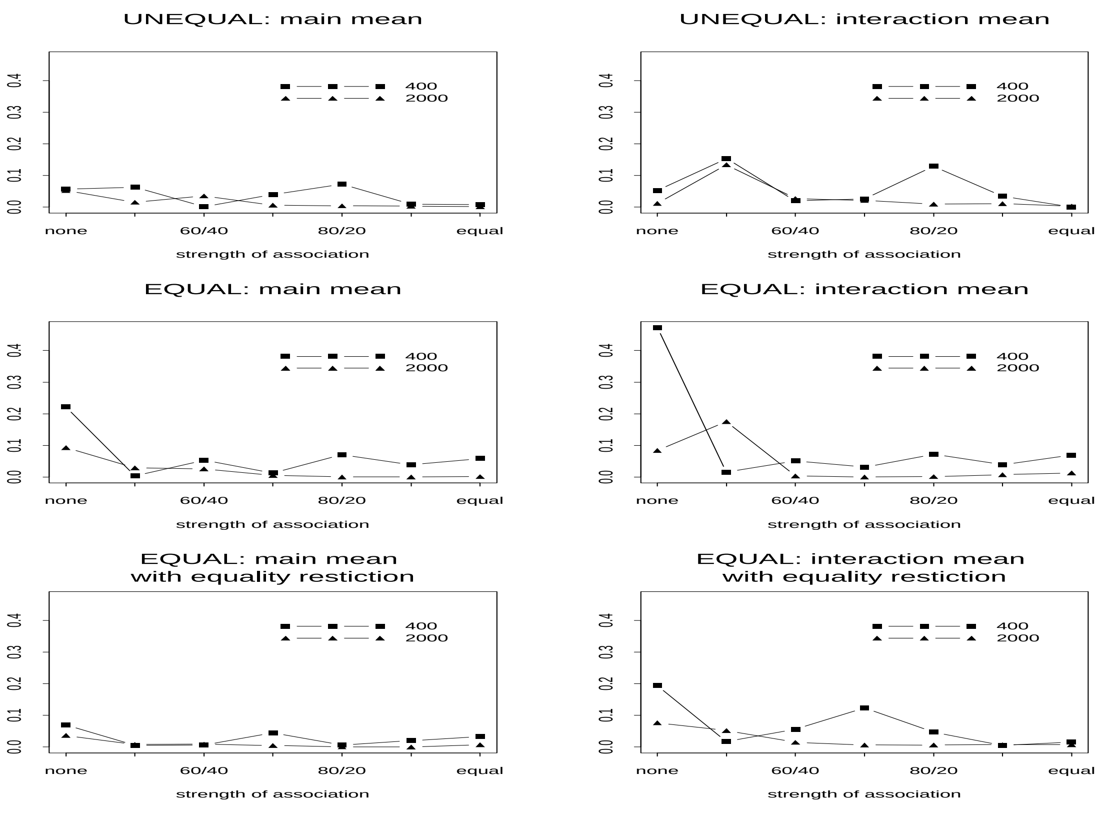

Figure 1: Mean deviances of the difficulty parameters. Obtained by aggregating over all items and replications for every –association × number of subjects × parameter set– combination.

The absolute value of the mean deviance is plotted against the strength of association between the collateral variable and the latent class. The absolute value is plotted to facilitate comparisons. Figure 1 shows that with four hundred subjects, the deviances, on average, don't systematically become smaller as stronger collateral information is incorporated. On the other hand, for data sets with two thousand subjects, the deviances seem to reduce as stronger collateral variables are incorporated, and the mean is close to zero from roughly the "60/40" condition onwards.

For the discrimination parameters the results are plotted in Figure 2.

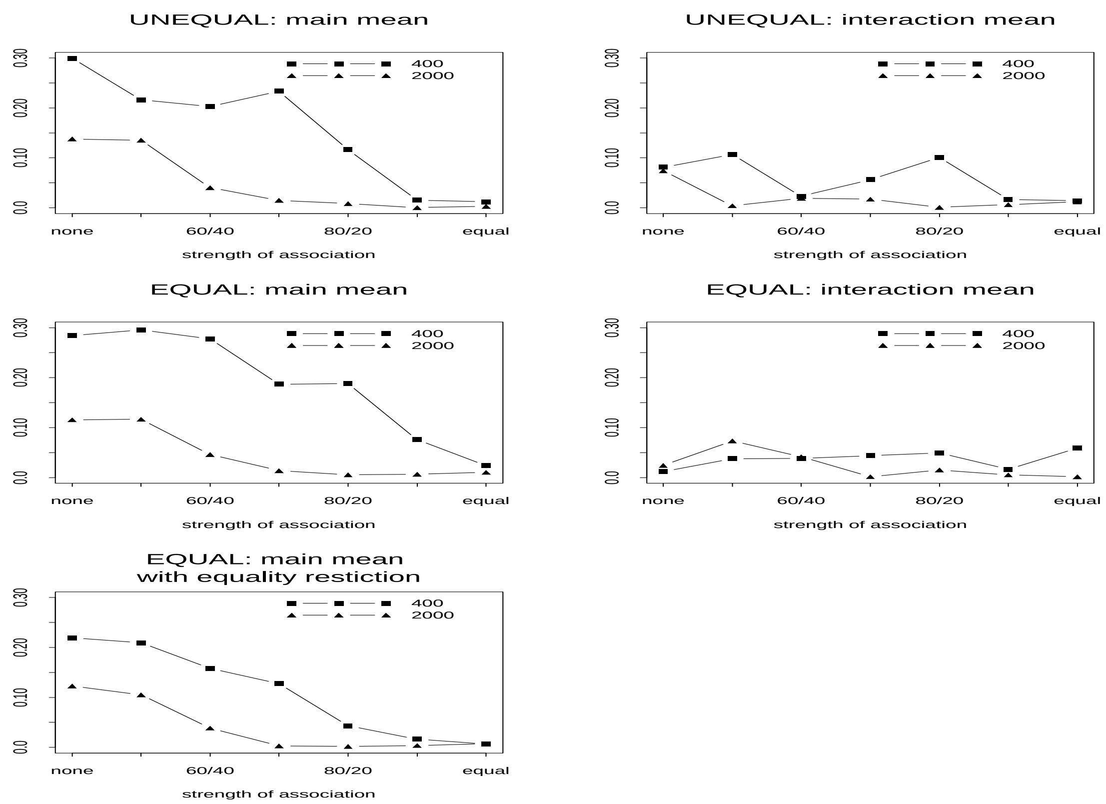

Figure 2: Mean deviances of discrimination parameters. Obtained by aggregating over all items and replications for every –association × number of subjects × parameter set–combination.

From this figure it can be seen that on average the estimated parameters are closer to their true values when the collateral information becomes stronger. However the difference in item discrimination in both latent classes (the item × latent class interaction) does not seem to improve systematically. Moreover, both the discrimination and difficulty parameters have smaller mean deviances in data sets with two thousand than four hundred subjects.

4.2 Standard deviation of the deviances

Next we evaluate the variability of the estimated item parameters. It is expected that the deviances reduce as stronger collateral information is incorporated into the analysis. From the *missing information principle* ((Orchard & Woodbury, 1972)) we know that the complete information is composed of the observed information plus the missing information. By incorporating a variable that associates with the missing variable (latent class), the observed information increases, thus reducing the variability of the parameter estimates. First, the standard deviation of the deviances of the difficulty parameters is plotted in Figure 3.

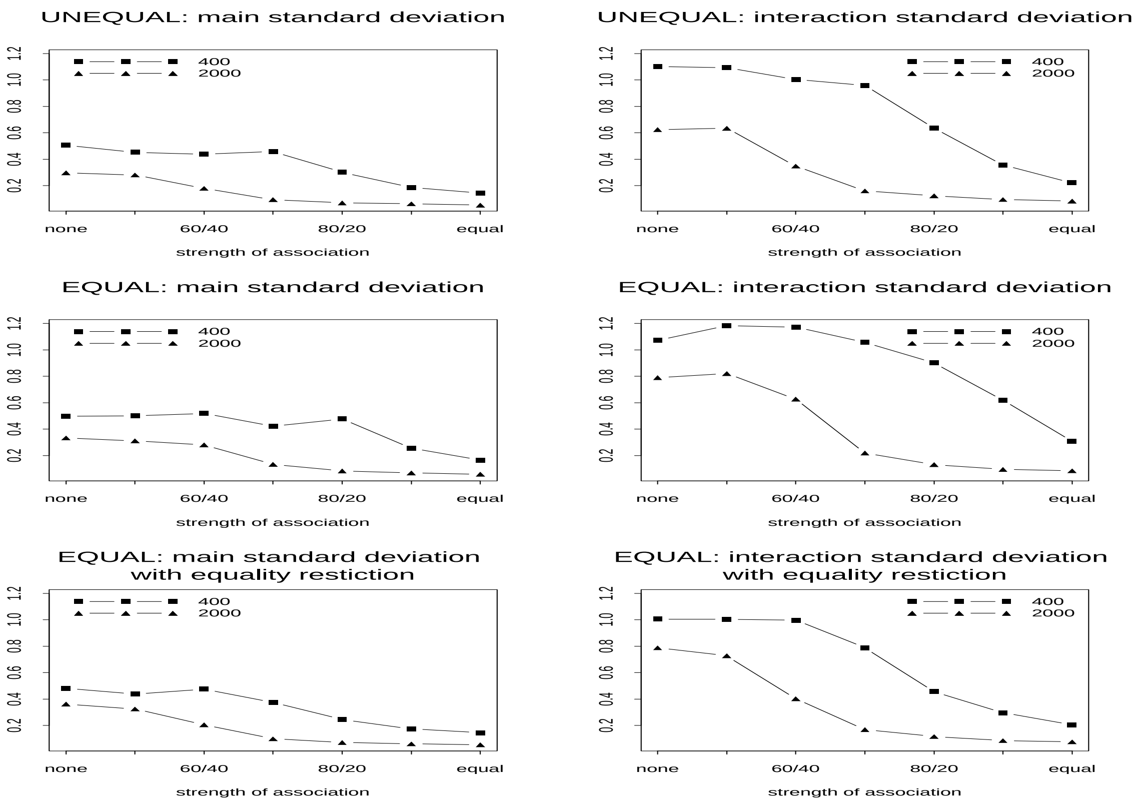

Figure 3: Standard deviations of deviances of the difficulty parameters. Obtained by aggregating over all items and replications for every –association × number of subjects × parameter set– combination.

The standard deviations reduce systematically as the association between the latent class and the collateral variable becomes stronger. In an earlier paper (Smit et al., 1999) the Mixed Rasch model (estimated via CML) was evaluated with a similar simulation study. For data sets with two thousand subjects, the standard errors of the deviances found were very similar with the standard errors reported in Figure 3. In Figure 4 the standard deviations of the discrimination parameters are plotted.

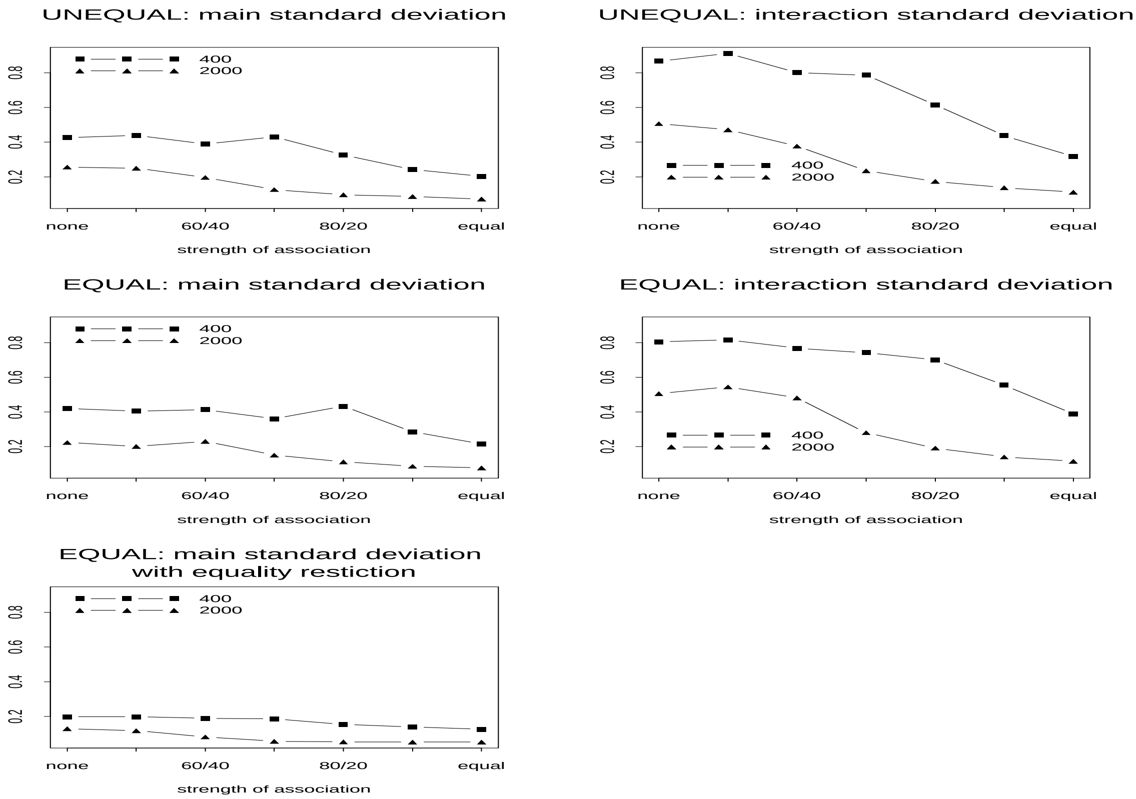

Figure 4: Standard deviation of deviances of the discrimination parameters. Obtained by aggregating over all items and replications for every –association × number of subjects × parameter set– combination.

Again, the main trend is that standard deviations reduce as collateral information is more strongly associated with the latent class. Furthermore, the equality restriction on the discrimination parameters more or less halves the standard deviations.

4.3 Iterations

In Figure 5 the 25^{th}, 50^{th}, and 75^{th} percentile of the number of iterations needed to converge are reported.

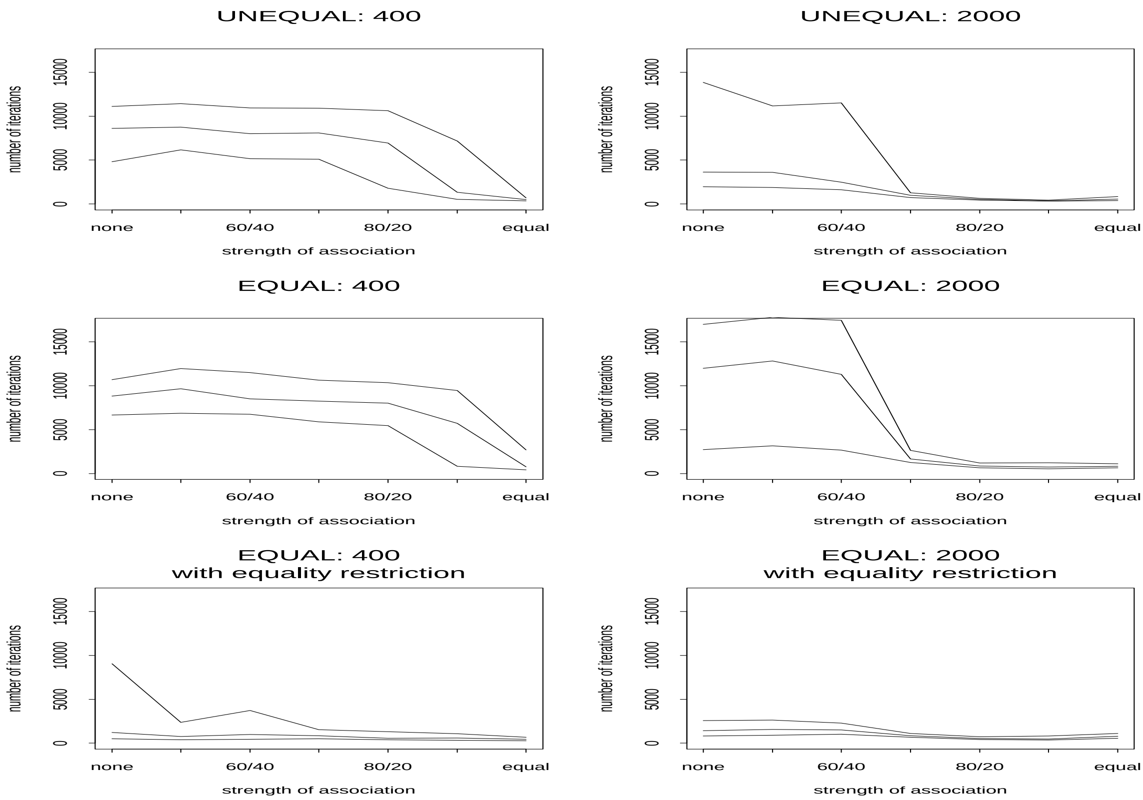

Figure 5: 25^{th}, 50^{th} and 75^{th} percentile of iterations needed to converge, aggregating over all items and replications within every –association $\times$ number of subjects $\times$ parameter set–combination.

With two thousand subjects, the equality restriction on the discrimination parameters is very effective when little collateral information is used. At the "70/30" condition the number of iterations sharply decreases when no restriction is used. With four hundred subjects, on the other hand, the equality restriction seems to be the only factor to seriously speed up convergence, the strength of the collateral variable doesn't seem to have a strong effect.

4.4 Latent class assignment

Finally the potential to assign subjects to the latent classes is evaluated. For every response pattern, posterior probabilities can be computed for the latent classes. The ratio of these probabilities can be used to assign all subjects to one of the latent classes. The classification is done using the item pattern only, and using both the item pattern and the collateral variable.

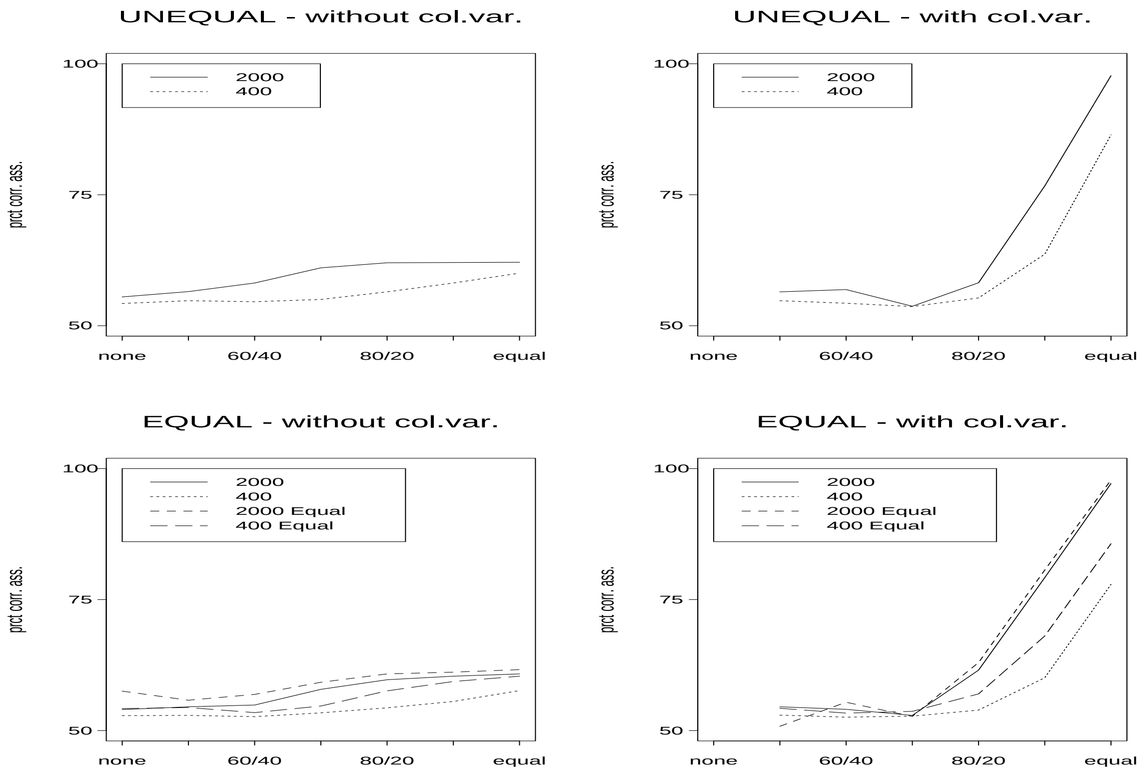

Figure 6: Percentage correctly assigned subjects.

The percentage correctly classified subjects is bounded when using the items only. This is because each item pattern is assigned to one latent class, while it is perfectly feasible for two subjects from different latent classes to have the same item pattern. The left figure shows that the bound is approached if the number of subjects increases, or if the association between the latent class and the collateral variable becomes stronger (due to better parameter estimates). The right figure shows that as the collateral information becomes better the posterior probabilities get dominated by this variable, and thus approach high percentages of correctly classified subjects.

5 Conclusion

The incorporation of collateral information into the model, reduces the standard errors in the item parameter estimates considerably. Fitting a Birnbaum model in small samples can in itself be problematic. So, fitting a Mixed Birnbaum model with four hundred cases is bound to be troublesome, unless the classes are expected to answer according to widely different models (large differences in item parameters between the latent classes). In most practical situations it seems advisable to obtain as much collateral information as possible before applying the Mixed Birnbaum model.

We would like to note that the proposed model can easily be extended. For instance, generalizations to polytomuous collateral variables or more latent classes are trivial. Since multiple categorical variables can be represented as one joint categorical variable the incorporation of multiple collateral variables is also trivial. Furthermore, an unsaturated logit model for the latent class variable on the collateral variables can be specified, enabling continuous collateral variables.

Finally, the latent class variable itself can be polytomuous. Note however that the number of item parameters grows rapidly if no restrictions are imposed.

Within the proposed framework, there seems to be ample possibilities to stabilize the model by incorporating collateral variables. More research into this area is needed to enlarge the applicability of the Mixed Birnbaum model in practical situations. But the results obtained in this simulation study seem promising.

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (p. 397-479). Reading, MA: MIT Press.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal Royal Statistical Society Series B*, *39*(1), 1–38. (With discussion)

Fisher, G. H., & Molenaar, I. W. (1995). *Rasch models - foundations, recent developments, and applications.* New York: Springer.

Guttman, L. (1944). A basis for scaling qualitative data. *American-Sociological-Review*, *9*, 139-150.

Hagenaars, J. A. (1993). *Loglinear models with latent variables.* Sage Publications.

Heinen, T. (1993). *Discrete latent variable models.* Tilburg: University Press.

Kelderman, H., & Macready, G. B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement*, *27*(4), 307–327.

Langeheine, R., & Rost, J. (1988). *Latent trait and latent class models.* New York: Plenum.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Linden, W. Van der, & Hambleton, R. (1997). *Handbook of modern item response theory.* Berlin: Springer.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*(2), 195 – 215.

Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. In L. M. Le Cam, J. Neyman, & E. L. Scott (Eds.), *Proceedings of the sixth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 697–715). University of California Press.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhage Nielsen & Lydicke.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*, 271-282.

Rost, J., & Langeheine, R. (1997). *Applications of latent trait and latent class models in the social siences.* Waxmann Verlag GmbH.

Samejima, F. (1969). Estimation of latent ability using a pattern of graded scores. *Psychometric Monograph*(17).

Smit, A., Kelderman, H., & van der Flier, H. (1999). Collateral information and mixed rasch models. *Methods of Psychological Research*, *4*(3), 19–32.

Vermunt, J. (1997). *Log-linear models for event histories.* Thousand Oaks, CA, USA: Sage Publications, Inc.

Wilson, M., Draney, K., & Engelhard, J., G. (1999). *Objective measurement: Theory into practice* (Vol. 5). Norwood, NJ: Ablex.

Wilson, M., Engelhard Jr., G., & Draney, K. (1997). *Objective measurement: Theory into practice* (Vol. 4). Norwood, NJ: Ablex.