Published in final edited form as:

Comput Stat Data Anal. 2013 January ; 57(1): 392–403. doi:10.1016/j.csda.2012.06.022.

Limited Information Estimation in Binary Factor Analysis: A Review and Extension

Jianmin Wu^a and Peter M. Bentler^b

^aBeijing Institute of Technology

^bUniversity of California, Los Angeles

Abstract

Based on the Bayes modal estimate of factor scores in binary latent variable models, this paper proposes two new limited information estimators for the factor analysis model with a logistic link function for binary data based on Bernoulli distributions up to the second and the third order with maximum likelihood estimation and Laplace approximations to required integrals. These estimators and two existing limited information weighted least squares estimators are studied empirically. The limited information estimators compare favorably to full information estimators based on marginal maximum likelihood, MCMC, and multinomial distribution with a Laplace approximation methodology. Among the various estimators, Maydeu-Olivares and Joe's (2005) weighted least squares limited information estimators implemented with Laplace approximations for probabilities are shown in a simulation to have the best root mean square errors.

Keywords

Limited Information; Laplace Approximation; Binary Response; Marginal Likelihood; Factor Scores

1. Introduction

Factor analysis is a very popular and well-developed model for the analysis of continuous data (see e.g., Yanai and Ichikawa, 2007). While the factor analysis model is equally applicable to binary response data, the development of a reliable statistical machinery for it has resisted consensus. An ideal class of methods involves the use of full information maximum likelihood. However, with the inevitable sparseness of data for models with many variables and dimensions and the difficult integrations involved, limited information methods recently have been revived as a serious contender for routine use.

The development of marginal maximum likelihood (MML) estimation of factor loadings provided an important boost to full information methods (Bock and Aitkin, 1981, Meng and Schilling, 1996). Maximum marginal likelihood estimation is, however, often complicated because the marginal likelihood includes an intractable integral in estimation of categorical data models in factor analysis. Several approaches have been developed to treat this troublesome integral. Numerical integration is one option, for example, Naylor and Smith

© 2012 Elsevier B.V. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(1982) used Gaussian quadrature which leads to the efficient calculation of posterior densities. In numerical analysis, Gauss-Hermite (G-H) quadrature is an extension of the Gaussian quadrature method for approximating the value of integrals. A main problem with G-H quadrature is the inaccurate approximation of integrals. While adaptive G-H quadrature (Pinheiro and Chao, 2006, Schilling and Bock, 2005) can improve accuracy, it also is not optimal when the dimensionality of the integral is high because the number of quadrature points can grow exponentially with the number of latent variables. A second approach involves the use of EM and MCEM algorithms to fit various latent variable models (Meng and van Dyk, 1998, 1997). They have advantages of numerical stability and implementation simplicity, but also have two disadvantages. They are often criticized for slow convergence when the fraction of missing information is large. Also, the computational efficiencies of MCEM algorithms (An and Bentler, 2012, Chan and Kuk, 1997, Levine and Casella, 2001) rely heavily on the convergence rate of corresponding MCMC samplers. Still another approach involves Laplace approximation, which uses a large sample normal approximation to the posterior distribution and is generally very accurate. Applications of this approach can be found in Breslow and Clayton (1993), Lee and Nelder (1996, 2006), Lin and Breslow (1996), and Huber et al. (2004). Laplace approximation maximum likelihood estimation (LAML) has been widely used for integrals in Bayesian inference (Tierney and Kadane, 1986). The Laplace approximation is also useful for approximating the likelihood in various nonlinear random effects models, when the integrals in the likelihood do not have closed form solutions. Liu and Pierce (1993) showed that Laplace approximation to the marginal likelihood is very accurate. Yun and Lee (2004) implemented the h-likelihood based Laplace approximation for fixed effects, and compared it to G-H for binary outcomes. Raudenbush et al. (2000) applied a Laplace approximation to the integral and maximized the approximate integrated likelihood via Fisher scoring. On the other hand, Joe (2008) reported that the Laplace approximation is biased in generalized linear mixed models with binary response variables. Thus, it is not known whether the Laplace method can be used in binary factor analysis models. This paper reviews LAML in this context (Wu and Bentler, 2012) as well as an alternative method based on the multinomial distribution, and also utilizes MML and MCMC methods to compare to the main methods of interest, limited information methods based on the Bernoulli and multinomial distributions.

Limited information methods were developed to avoid the problems of high dimensional normal integrals, as well as to address the problem that, due to the consequent sparseness of contingency tables, estimation of parameters in a multinomial distribution framework becomes increasingly difficult as the number of multinomial categories increases (e.g., Agresti, 2002, sec. 9.8). Although these methods have a long history (e.g., Christofferson, 1975), in recent years Maydeu-Olivares (Forero and Maydeu-Olivares, 2009, Lee et al., 1995, Maydeu-Olivares, 2001, Maydeu-Olivares and Joe, 2005, Maydeu-Olivares, 2006) has been the main proponent for the use of limited information methods to estimate latent trait models based on univariate and bivariate Bernoulli sample moments with weighted least squares. See also (Bartholomew and Knott, 1999, Christofferson, 1975). More recently, in a longitudinal modeling framework Fu et al. (2011) showed that a pairwise likelihood approach with an EM algorithm and quadrature also can perform comparably to full information methods, but they do not compare their method to other limited information methods. As far as we know, outside of their developers, there has been no systematic evaluation of the proposed limited information weighted least squares estimator. We study these estimators in factor analysis models where they have not yet been applied with Laplace approximations to joint probabilities. Furthermore, we develop an alternative limited information estimator based on the first and second order marginal moment maximum likelihood, using marginal moments obtained by mapping between marginal moment and joint probability as approximated by the Laplace method and an extended likelihood or Bayes model approach to factor scores as estimated by Newton-Raphson. We compare this

new approach to limited information estimation with that of weighted least squares as well as with full information methods.

In the context of latent trait and factor analysis models, an important issue is testing the adequacy of any proposed model. Thus we also report goodness of model fit using limited information by using a quadratic form statistic that is asymptotically chi-square distributed (Christofferson, 1975, Maydeu-Olivares and Joe, 2005, Reiser, 1996).

The paper is organized as follows: in section 2, we introduce a model for factor analysis with binary response data. In section 3 we discuss factor score and factor loading estimation. In section 4, we review limited information test statistics of model fit. In section 5 we provide a comparison of estimates of several methods, and report on simulations that study their performance. section 6 gives the conclusions.

2. A Model for Factor Analysis

In this section we introduce notation for a factor analysis model with a logistic link function for binary response variables. These observed or manifest variables are denoted by $Y_j = (y_{j1}, y_{j2}, \dots, y_{jc})^T$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, c$. y_{ij} is i th independent response from the j th item, and $y_{ij} = \begin{cases} 1 & \text{success} \\ 0 & \text{others (or: yes vs. no, agree vs. disagree, etc.)} \end{cases}$. A latent structure model supposes the binary variables are related to a set of k unobservable variables denoted by $\eta_i = (\eta_{i1}, \eta_{i2}, \dots, \eta_{ik})^T$, where η_i denotes a $k \times 1$ vector of factor scores from i th response. For simplicity of argument, as is usually assumed in exploratory factor analysis, the latent variables are assumed to follow a multivariate normal distribution with a null mean vector $E(\eta_i) = 0$ and identity covariance matrix $\text{cov}(\eta_i) = I_{k \times k}$. Letting $\Lambda_j = (\lambda_{1j}, \lambda_{2j}, \dots, \lambda_{kj})^T$ be a $k \times 1$ vector and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_c)$ be a $1 \times c$ intercept vector, conditioning on the parameters α , Λ and η , the conditional probability of success is given by

$$\pi_{ij} = P(y_{ij} | \alpha_j, \eta_i, \Lambda_j) = \frac{\exp(\alpha_j + \eta_i^T \Lambda_j)}{1 + \exp(\alpha_j + \eta_i^T \Lambda_j)} \quad (1)$$

3. Estimation of Factor Scores and Factor Loadings

In standard applications of MML in item response theory (IRT), latent trait scores do not impact estimates of the item parameters. They are computed once, conditional on estimates of the parameters. In our approach to the factor analysis model with binary response data, we consider both the factor loadings and latent factor scores as parameters to be estimated. That is, we allow the factor score estimates to impact the item parameters, but we do so with an extended likelihood approach to avoid treating the factor scores as fixed parameters that are equivalent to item parameters. Computationally, we do this sequentially, first estimating the factor scores for fixed factor loadings, and second, estimating the factor loadings given that factor scores are fixed.

3.1. Estimation of factor scores

We use the extended likelihood to estimate factor scores in the binary factor analysis model. Extended likelihood for random variables is a methodology that has existed for quite some time, see e.g., Butler (1986) and Bjornstad (1996). Recently, Lee et al. (2006) gave a detailed justification for extended likelihood. The idea of extended likelihood is essentially the same as that of the Bayes modal estimator that is used in item response theory (e.g.,

Bock and Aitkin, 1981). The Bayes modal estimator is based on the maximum of the posterior distribution density

$$f_{\alpha,\Lambda}(\eta|Y) = \frac{\pi_{\alpha,\Lambda}(Y|\eta) f_{\alpha,\Lambda}(\eta)}{\int \pi_{\alpha,\Lambda}(Y|\eta) f_{\alpha,\Lambda}(\eta) d\eta} \quad (2)$$

so the $f_{\alpha,\Lambda}(\eta|Y) \propto \pi_{\alpha,\Lambda}(Y|\eta) f_{\alpha,\Lambda}(\eta)$. Thus we consider

$$L_i = \pi(y_i|\eta_i) f(\eta_i), i=1, 2, \dots, N \quad (3)$$

Taking logs of both sides of equation (1) and using (3) we obtain the log likelihood function

$$\ell_i = \sum_{j=1}^c \left[y_{ij} \ln \pi_{ij} + (1 - y_{ij}) \ln (1 - \pi_{ij}) \right] - \frac{k}{2} \ln (2\pi) - \frac{1}{2} \eta_i^T \eta_i, i=1, 2, \dots, N \quad (4)$$

Using this joint likelihood, we estimate the factor scores with the profile likelihood of $\alpha\Lambda$, as given by

$$\ell_i(\alpha, \Lambda) = \max_{\eta_i} \ell_i(\alpha, \Lambda, \eta_i), i=1, 2, \dots, N \quad (5)$$

In (5), maximization is performed at a fixed value of α, Λ . To maximize (5) we use the Newton-Raphson algorithm based on first and second order derivatives with respect to η as given by

$$\frac{\partial \ell_i}{\partial \eta_i} = \sum_{j=1}^c y_{ij} \Lambda_j - \sum_{j=1}^c \pi_{ij} \Lambda_j - \eta_i, i=1, 2, \dots, N \quad (6)$$

$$\frac{\partial^2 \ell_i}{\partial \eta_i \partial \eta_i^T} = - \left\{ \sum_{j=1}^c \left[\pi_{ij} (1 - \pi_{ij}) \Lambda_j \Lambda_j^T \right] + I_k \right\}, i=1, 2, \dots, N \quad (7)$$

Then, a step of the iterative Newton-Raphson algorithm involves computing

$$\eta_i^{(s+1)} = \eta_i^{(s)} - \left[\ell_i^{(2)}(\eta_i^{(s)}) \right]^{-1} \ell_i^{(1)}(\eta_i^{(s)}), i=1, 2, \dots, N \quad (8)$$

This method can break down if (7) is not positive definite, and stephalving may be required if (8) does not improve the function value. We did not encounter these problems in our computations.

3.2. Estimation of factor loadings

In this section we review and develop several estimators: LIWLS, based on limited information weighted least squares; FIMDML, based on full information multinomial distribution maximum likelihood; LIMDML, based on limited information multinomial distribution maximum likelihood; and LAML, based on Laplace approximation maximum likelihood. LIWLS and LIMDML are given in two versions with an added last letter “2” or “3” to designate the use of up to 2nd order or up to 2rd order moment information.

3.2.1. Estimation of factor loadings with limited information—Consider one of the 2^c response patterns $Y_i = (y_{i1}, y_{i2}, \dots, y_{ic})$, a random vector that is a c dimensional Bernoulli

random variable, $y_{ij} = \begin{cases} 1 & \text{success} \\ 0 & \text{others} \end{cases}$, $i = 1, 2, \dots, 2^c$, based on probabilities

$$\pi_i = p[Y_i = (y_{i1}, y_{i2}, \dots, y_{ic})], i = 1, 2, \dots, 2^c \quad (9)$$

The marginal probabilities for the binary model are computed using the Laplace approximation. Details of the calculating process is given in Appendix A.1. Without loss of generality, the elements of π are ordered by lexicographic ordering of the response patterns in π (Cai et al., 2006, Maydeu-Olivares and Joe, 2005). Let N denote the number of independent trials and the random variable n_i be the number of times outcome Y_i is observed

over the N trials, $\sum_{i=1}^{2^c} n_i = N$, where each trial results in exactly one of some fixed finite number 2^c of possible outcomes, with joint probabilities $\pi = (\pi_1, \pi_2, \dots, \pi_{2^c})^T$, $\sum_{i=1}^{2^c} \pi_i = 1$. The probability function of the multinomial distribution is given (e.g., Agresti, 2002, p.6) by:

$$p(Y_1 = n_1, Y_2 = n_2, \dots, Y_{2^c} = n_{2^c}) = \frac{N!}{n_1! \dots n_{2^c}!} \pi_1^{n_1} \dots \pi_{2^c}^{n_{2^c}} \quad (10)$$

Next we review known structural relationships between joint moments and joint probabilities.

Let $\dot{\pi}_j$ express the C_c^j dimensional vector with elements

$$\pi_{jd} = E\left\{\overbrace{y_s \dots y_k \dots y_t}^{s < k < t}\right\}_j = p(y_s = 1, \dots, y_k = 1, \dots, y_t = 1), \quad (11)$$

where $j = 1, 2, \dots, c$, $d = 1, 2, \dots, C_c^j$. Then the relationship between joint moments and joint probabilities of the multinomial distribution can be described as follows (Maydeu-Olivares and Joe, 2005, Teugels, 1990). The c dimensional random vector $Y_i = (y_{i1}, y_{i2}, \dots, y_{ic})$ of Bernoulli distribution can be characterized by the $2^c - 1$ dimensional vector $\dot{\pi}$ of its joint moments. There is a $T_{(2^c-1) \times 2^c}$ matrix that satisfies

$$\dot{\pi} = T\pi \quad (12)$$

Formula (12) is used to transform joint probabilities to joint moments. Let p and $\dot{p}$ represent the 2^c dimensional vector of cell proportions and the $2^c - 1$ dimensional vector of sample joint moments, respectively. A relationship exists between residuals of the sample joint and population moments, as well as the residuals of sample proportions and population probabilities, see e.g. (Agresti, 2002, p.576-594), namely

$$\sqrt{N}(\dot{p} - \dot{\pi}) = \sqrt{N}T(p - \pi) \quad (13)$$

It is also well known that

$$\sqrt{NT} (p - \pi) \xrightarrow{d} N(0, \Gamma), \quad (14)$$

and hence the discrepancy between sample and population moments converge in distribution to a multivariate normal distribution based on the multivariate central limit theorem (Rao, 1973, p.128) with

$$\sqrt{N} (\hat{p} - \pi) \xrightarrow{d} N(0, \Xi), \quad (15)$$

where $\Xi = \Pi \Gamma \Pi^T$, $\Gamma = D - \pi \pi^T$, and $D = \text{diag}(\pi)$.

Let p_r denote the vector of sample moments up to order r , with dimension $\sum_{j=1}^r C_c^j$. Then the r order population moment can be written as $\pi_r = T_r \pi$.

In order to get the limited information weighted least squares estimator—denoted generically as LIWLS and specifically as LIWLS2 and LIWLS3, depending on whether $r=2$ or $r=3$, we minimize $L_r = N(p_r - \pi_r)^T \Xi_r^{-1} (p_r - \pi_r)$ with $\theta = (\alpha, \Lambda)^T$ to obtain

$$\hat{\theta} = \arg \min_{\theta} L_r(\theta) \quad (16)$$

At the same time, when $\hat{\theta}$ is the maximum likelihood estimator (MLE) or another consistent minimum variance estimator, Bishop et al. (1975, p. 457-530) showed that

$$\sqrt{N} (\hat{\theta} - \theta) = B \sqrt{N} (p - \pi(\theta)) + o_p(1) \quad (17)$$

where $B = I^{-1} \Delta^T D^{-1} \Delta$ with the consequence that

$$\sqrt{N} (\hat{\theta} - \theta) \xrightarrow{d} N(0, I^{-1}), \quad (18)$$

where $I = \Delta^T D^{-1} \Delta$ is the Fisher information matrix. This formula gives variance of the estimator that is used below. we now have all the details needed for our algorithm to estimate the factor scores as well as the factor loadings. Its steps are:

1. Start with an estimate of the factor loadings;
2. Using (8) and given the current estimate of $\hat{\eta}$, update $\hat{\theta}$ by solving (16);
3. Given the current matrix of $\hat{\theta}$, update $\hat{\eta}$ by solving (8);
4. Iterate between steps 1 and 3 until convergence.

3.2.2. Limited information multinomial MLE approach to factor loadings

estimation—Given the probability function of the multinomial distribution (see eq. 10), the log likelihood kernel function of the multinomial probability distribution can be written as follows:

$$\ell(\theta) \propto \sum_{i=1}^{2^c} n_i \log \pi_i(\theta) \quad (19)$$

(see, e.g., Agresti, 2002, p.21). We may call an estimator that maximizes this function the full information multinomial distribution maximum likelihood estimator (FIMDML). It can

be obtained by using (19) in step 2 in the algorithms of this section. Although this is an obvious estimator, it seems not to have been used in binary factor analysis with logistic link functions. However, as noted, with sparse contingency tables it is not a good idea to use (19) directly. Instead, we just use one and two dimensional Bernoulli distributions to describe structural relationship between manifest variables. To do this, we need to obtain the first moments and the 2nd order joint moments.

With random vector $Y_i = (y_{i1}, y_{i2}, \dots, y_{ic})$ being a c dimensional Bernoulli random variable from the i th respondent,

$$p(y_i=1) = p_j, j=1, 2, \dots, c \quad (20)$$

$$p_{11}(i, j) = p(y_i=1, y_j=1) = p_{ij} \quad i \neq j, i < j, i, j=1, 2, \dots, c \quad (21)$$

From (20) and (21), the two dimensional Bernoulli distribution can be obtained.

Let $X_j, j=1, 2, \dots, c$ denote the number of times outcome number $y_j=1$ is observed over the N trials, then X_j follows a Bernoulli distribution, a special case of the multinomial distribution, with parameters α, Λ . Also, let $Z_{00}(i, j), Z_{10}(i, j), Z_{01}(i, j)$, and $Z_{11}(i, j)$ indicate the number of times the outcome (y_{ti}, y_{tj}) is observed, representing the vector $(0, 0), (1, 0), (0, 1), (1, 1)$, respectively, t over the N trials. Then the vector $(Z_{00}(i, j), Z_{10}(i, j), Z_{01}(i, j), Z_{11}(i, j))$ also follows a multinomial distribution with parameters θ . The likelihood kernel function is constructed as:

$$L_1(\theta) \propto \prod_{i=1}^c p_i^{X_i} (1 - p_i)^{(N - X_i)}, \quad (22)$$

$$L_2(\theta) \propto \prod_{i < j, i \neq j, i=1}^c p_{00}(i, j)^{Z_{00}(i, j)} p_{10}(i, j)^{Z_{10}(i, j)} p_{01}(i, j)^{Z_{01}(i, j)} p_{11}(i, j)^{Z_{11}(i, j)}, \quad (23)$$

$$\ell_2(\theta) \propto \log(L_1(\theta) L_2(\theta)), \quad (24)$$

Finally, we have the required elements to define the LIMDML estimator

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} [\ell_2(\theta)], \quad (25)$$

which we denote as LIMDML2 since it is based on the first two moments. As before, we now have all the ingredients for our algorithm to estimate factor scores and factor loadings:

1. Start with an estimate of the factor loadings;
2. Using (8) and given the current estimate of $\widehat{\eta}$, update $\widehat{\theta}$ by solving (25);
3. Given the current matrix of $\widehat{\theta}$, update $\widehat{\eta}$ by solving (8);
4. Iterate between step 1 and 3 until convergence.

Expanding this methodology to allow the use of more information may provide more accurate estimators. Using a similar notation as above,

$$L_3(\theta) \propto \prod_{i < j < k}^c P_{000}^{z_{000}} P_{100}^{z_{100}} P_{010}^{z_{010}} P_{001}^{z_{001}} P_{110}^{z_{110}} P_{101}^{z_{101}} P_{011}^{z_{011}} P_{111}^{z_{111}} \quad (26)$$

$$\ell_3(\theta) \propto \log(L_1(\theta) L_2(\theta) L_3(\theta)), \quad (27)$$

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} [\ell_3(\theta)], \quad (28)$$

In parallel to the previous notation, we denote this LIMDML estimator as LIMDML3 since it is based on the first three moments. We also have the following algorithm.

1. Start with an estimate of the factor loadings;
2. Using (8) and given the current estimate of $\widehat{\eta}$, update $\widehat{\theta}$ by the solving (28);
3. Given the current matrix of $\widehat{\theta}$, update $\widehat{\eta}$ by solving (8);
4. Iterate between steps 1 and 3 until convergence.

3.2.3. Estimation of factor loadings using Laplace approximation—In addition to using the Laplace approximation for limited information, for coherence and comparison in performance, we also summarize the approach of Wu and Bentler (2012) for full information estimation of factor loadings. As before, with $Y_i = (y_{i1}, y_{i2}, \dots, y_{ic})$ a c

dimensional Bernoulli random variable with $y_{ij} = \begin{cases} 1 & \text{success} \\ 0 & \text{others} \end{cases}$, $i=1, 2, \dots, N$, and using our logistic model (1), we obtain the marginal likelihood function by

$$L = \prod_{i=1}^N \int \exp(\ell_i) d\eta_i \quad (29)$$

Where $\ell_i = \sum_{j=1}^c [y_{ij} \ln \pi_{ij} + (1 - y_{ij}) \ln (1 - \pi_{ij})] - \frac{k}{2} \ln 2\pi - \frac{1}{2} \eta_i^T \eta_i$. So the *log* marginal likelihood function is given by

$$\ell(\theta) = \sum_{i=1}^N \log \left(\int \exp(\ell_i) d\eta_i \right) \quad (30)$$

Solving the intractable integral above is difficult due to the complexity of the nonlinear integrand. The marginal *log* likelihood function ℓ is given approximately using the Laplace method. From Appendix A.1, we know

$$\ell(\theta) = \sum_{i=1}^N \log \pi_i(\theta) \quad (31)$$

where $\log(\pi_i(\theta)) = \ell_i(\theta) + \frac{k}{2} \ln 2\pi + \frac{1}{2} \ln \left| \sum_i \cdot \right|$. So we just need to maximize the adjusted profile likelihood in θ , for writing convenience, according to the notation above in 3.2.1,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} [\ell(\theta)], \quad (32)$$

Wu and Bentler call this estimator LAML, the Laplace approximation maximum likelihood. They also obtain the following algorithm, which is similar in structure to the previous ones.

1. Start with an estimate of the factor loadings;
2. Using (8) and given the current estimate of $\hat{\eta}$, update $\hat{\theta}$ by solving (32);
3. Given the current matrix of $\hat{\theta}$, update $\hat{\eta}$ by solving (8);
4. Iterate between step 1 and 3 until convergence.

Some additional details on the LAML methodology are given in J. Wu and Bentler (2012).

4. The limited information test statistics

In this section we summarize the goodness of fit chi-squared statistics that are relevant to testing the adequacy of our binary factor analysis model. We provide the relevant known the test statistics that we can be used to evaluate simple and composite null hypotheses based on both full and limited information.

As given previously, with $Y_i = (y_{i1}, y_{i2}, \dots, y_{ic})^T$ the i th pattern, we have cell counts $n_1, n_2, \dots, n_{2^c}$ that have a multinomial distribution with cell probabilities $\pi = (\pi_1, \pi_2, \dots, \pi_{2^c})^T$. So $N = n_1 + n_2 + \dots + n_{2^c}$ and $p = (p_1, p_2, \dots, p_{2^c})^T$ denotes the sample proportions, where $p_i = n_i/N$. Then, as is well-known, Pearson's test statistic for a simple null hypothesis $H_0: \pi = \pi_0$ is

$$\chi^2 = N \sum_{i=1}^{2^c} \frac{(p_i - \pi_i)^2}{\pi_i}, \quad (33)$$

which asymptotically approaches a chi-square distribution with $2^c - 1$ degrees of freedom (see e.g., Agresti, 2002, section 1.5.2). Maydeu-Olivares and Joe (2005) employed

$L_r = N(p_r - \pi_r)^T \Xi_r^{-1} (p_r - \pi_r)$ up to r order to test a simple null hypothesis, see (16). In our paper, we also use this statistic to test our model. Pearson's test can be used for a composite null hypothesis $H_0: \pi = \pi(\theta)$, and we use it with the maximum likelihood estimate $\hat{\theta}$

$$\hat{\chi}^2 = N \sum_{i=1}^{2^c} \frac{(p_i - \pi_i(\hat{\theta}))^2}{\pi_i(\hat{\theta})}, \quad (34)$$

The asymptotic null distribution of (34) is chi-square with $2^c - 1 - q$ degrees of freedom, where q is the number of estimated parameters. In addition, as noted above we could use the asymptotic chi-squared tests

$$L_r(\theta) = N(p_r - \pi_r(\theta))^T \Xi_r^{-1}(\theta) (p_r - \pi_r(\theta)), \quad (35)$$

with $\hat{\pi}_r = \pi_r(\hat{\theta})$ and $\hat{\Xi}_r = \Xi_r(\hat{\theta})$, based on $s = \sum_{j=1}^r C_c^j - q$ degrees of freedom. However, Ξ_r is evidently not stable, so we also use the adjusted covariance matrix (Joe and Maydeu-Olivares, 2010, Khatri, 1966, Maydeu-Olivares and Joe, 2005)

$$C_r(\theta) = \Xi_r^{-1}(\theta) - \Xi_r^{-1}(\theta) \Delta_r(\theta) \left(\Delta_r^T(\theta) \Xi_r^{-1}(\theta) \Delta_r(\theta) \right)^{-1} \Delta_r^T(\theta) \Xi_r^{-1}(\theta), \quad (36)$$

$$M_r(\theta) = N(p_r - \pi_r(\theta))^T C_r(\theta) (p_r - \pi_r(\theta)), \quad (37)$$

where the $\Delta_r(\theta) = T_r \frac{\partial \pi(\theta)}{\partial \theta^T}$, with $s = \sum_{j=1}^r C_c^j - q$ degrees of freedom (Rao, 1973, p.364). Additional calculation details are given in Appendices A.1 and A.2.

5. An Example and Simulations

5.1. An example

The previously developed methods are studied in an example and with some small simulations in order to provide an idea of their performance in practice. We begin with a simulated binary factor analysis model based on 3000 subjects and 6 items for each of the 2 factors representing a simple structure. The data are generated according to the following logistic regression model

$$p(y_{ij}=1|\eta_i) = \frac{\exp(\alpha_j + \eta_i^T \Lambda_j)}{1 + \exp(\alpha_j + \eta_i^T \Lambda_j)} \quad (38)$$

where $\eta_i = (\eta_{i1}, \eta_{i2})$ follows a multivariate normal distribution with a null mean and identity covariance matrix and $\Lambda_j = (\lambda_{1j}, \lambda_{2j})^T$. For all estimators, iterations are terminated if and only if all differences in parameter estimates are less than a specified threshold, which is set to equal 1.0×10^{-4} in this paper. The algorithms for the methods described in section 3 were implemented using the MATLAB R 2009a programming language. MCMC and MML were obtained from EQSIRT (Wu and Bentler, 2011), as computed according to Sheng (2008) and Bock and Aitkin (1981) respectively. The results for the example are presented in Tables 1-4. The true values for the intercepts and factor loadings were selected randomly subject to some meaningful sign and size restrictions and are presented in the left-most data column in Table 1. This table also gives LIWLS estimators of order 2 and also of order 3, and their associated standard errors (SE) and SE averages, as well as root mean square errors (RMSE) between the true and estimated values for intercept parameters, factor loadings, and the overall combined performance. Table 1 shows that LIWLS 2 estimates of intercepts are marginally superior to those by LIWLS 3, while the reverse occurs for factor loadings. The average RMSE for all parameters is marginally better for LIWLS 3. Standard error estimates are about the same. Table 2 gives comparable results for LIMDML 2 and LIMDML 3, as well as those for FIMDML. With regard to the LIMDML estimates, as in LIWLS in the previous table, there is not much difference in precision between order 2 versus order 3, and also the standard errors are about the same. Comparing the two tables, LIMDML is more accurate than LIWLS in intercepts but worse in factor loadings. The last column of Table 2 gives results for FIMDML, which is more accurate than LIWLS and LIMDML in estimating intercepts, but LIWLS 3 beats FIMDML in accuracy for factor loadings. Standard error estimates are about the same.

Table 3 completes the results by giving the estimates for 3 full information methods, namely, MCMC, MML, and LAML. It can be seen that the MML and LAML estimates are about the same, and similar to those of FIMDML in Table 2, with all of these being better than those from MCMC. However, across all tables, the limited information method LIWLS 3 provides the most accurate estimates of factor loadings—even better than the full

information methods. Standard error estimates are not shown since they are all similar to those of FIMDML.

Table 4 provides some descriptive goodness of fit chi-squared tests along with p – values and degrees of freedom for all the methods. The left side gives the names of seven statistical approaches used to compute three different test statistics based on the associated estimator $\widehat{\theta}$. These statistics are M_2 and M_3 of eq. (37) and $\widehat{\chi}^2$ of eq. (34), and in each case, the test statistic, its p – value, and its degrees of freedom is presented. All tests lead to the same conclusion that the model is appropriate for the data. Perhaps the only interesting point is that the p – value for $\widehat{\chi}^2$, with its huge df, seems excessive.

5.2. Simulation Studies

The positive results for weighted least squares limited information estimators in the example served as a motivation for a small simulation of their behavior. The generating model was described in section 5.1 with two latent factors and 12 items. The sample size for both LIWLS 2 and LIWLS 3 is 2000. There are $rep = 50$ replications. Iterations terminated if and only if all differences in parameter estimates are less than 1.0×10^{-4} . The results, showing the population parameters and MSSE, defined as *mean of sum of square of errors* = $sum(estimate - true)^2 / replicated\ times$ are presented in Table 5.

These results show that LIWLS is a very promising method for evaluating binary factor analysis models. As in the example, accuracy of estimation is better with LIWLS 3 than with LIWLS 2, but the improvement is small. An additional simulation was done to evaluate the performance of the test statistics χ^2 given in (33) as well as $L_r = N(p_r - \pi_r)^T \Xi_r^{-1} (p_r - \pi_r)$ with $r = 2, 3$ to evaluate a simple null hypothesis. In this test, in order to distinguish parameter estimation and model testing, θ and hence π was taken as known, i.e., fixed at π_0 . The model was the same as before. Sample data were based on samples of size 1000. The test statistics were computed in each of 1000 replications, and the model was accepted or rejected at the 0.10 and 0.05 points in the chi-squared distribution. Table 6 gives the proportion of rejections of the true null hypothesis across these 1000 replications. If these tests performed ideally, they would reject the true model approximately 0.10 and 0.05 percent of the time.

The results show that all three statistics reject the true model too frequently, with χ^2 being the worst performing. Even though they tend to reject the true model too frequently, the limited information tests perform much more closely to nominal level than does χ^2 .

6. Conclusions and Remarks

We adapted Maydeu-Olivares and Joe (2005) existing limited information estimation method to factor analysis models with binary response variables using the extended likelihood or Bayes modal factor score estimator and the Laplace approximation to get required probabilities. We also developed some new alternatives including a full information method. A simulation demonstrated that limited information approaches are a promising strategy for parameter estimation in these models, and that Maydeu-Olivares and Joe's weighted least squares estimators are probably the most effective ones among the methods studied, with those based on information up to the 3rd-order being marginally better for estimation of factor loadings.

In contrast to Joe (2008), we found the Laplace method to perform adequately in estimating the parameters of the binary factor analysis model. In fact LIWLS 3 seemed to be even better than the full information MML method, with LAML also doing about as well as

MML. Of course, these results are preliminary and more thorough simulation work is needed to evaluate the generality of this observation.

Bock and Aitkin (1981) and others note that there are three main methodologies for computing or estimating latent trait scores. Historically, these take the model parameters as known, or fixed at the estimated parameter values, and compute factor score estimates conditional on these estimates. The first is the standard maximum likelihood estimator. It is not often used since it has the disadvantage of not giving finite score values for patterns with all correct or all incorrect responses. The second is the Bayes modal or maximum a posteriori (MAP), which has smaller variance compared with the former but shrinks estimates to the prior's mean. The third is Bayes EAP estimation (Bayes expected a posteriori), which has the advantage of simpler computations, see also (e.g., Baker, 2004, sec. 3.7). In contrast, our approach estimates the factor scores as part of the methodology to estimate factor loadings. As shown in section 3.1 and the appendix, our approach uses a profile likelihood maximization to obtain estimates of factor scores, conditional on values of intercepts and factor loadings. Since these Bayes modal factor scores are incidental to our main interest in limited information estimation, we did not study their properties specifically, and hence further work is needed to determine their behavior as compared to post-MML estimates of factor scores.

In the continuous case, it has long been known that linearly estimated factor scores based on estimated loadings are not equal to their true values and that the correlations among estimated scores do not accurately represent the actual factor correlations (e.g., McDonald and Burr, 1967). Although we use nonlinear estimation, this same phenomenon may well occur in the binary case. Because our algorithms update the intercepts and factor loadings using the currently estimated factor scores, it might be interesting to study whether making corrections for bias as is done in factor score regression (e.g., Skrondal and Laake, 2001, Hoshino and Bentler, 2012) could improve accuracy in estimation of factor loadings.

We studied test statistics that could be justified as being asymptotically chi-square distributed. The goodness of fit tests performed only adequately in the simulation, with models tending to be rejected too often, but the limited information tests performed better than the full information chi-square test as would be expected (e.g., Maydeu-Olivares and Joe, 2005). In the future, it may be useful to consider alternatives such as quadratic form statistics that are asymptotically distributed as a mixture of independent chi-square variates. Probabilities associated with such distributions can be directly computed (e.g., Satorra and Bentler, 1994) or else approximated with an alternative chi-square distribution obtained by matching moments (Bartholomew and Leung, 2002, Cai et al., 2006, Maydeu-Olivares, 2001, Satorra and Bentler, 1994, Yuan and Bentler, 2010). Although incidental to our main interest in limited information estimation, in an example the LAML and FIMDML full information estimators seemed to work about as well as those based on MML. However, a drawback of these approaches is that the speed of convergence appears to be impractically slow. Another disadvantage of FIMDML is that there is a computational limit since large matrices of joint probabilities need to be stored in high-dimensional problems. Further study to address these issues is warranted.

A.1 Estimation of marginal probability for binary models based on Laplace approximation

We obtain the marginal probability by $\pi_F P(Y_F(y_{i1}, y_{i2}, \dots, y_{ic})), i=1, 2, \dots, 2_c$

$$\pi_i(\alpha, \Lambda) = p(Y_i | \alpha, \Lambda, \eta_i) = \int \prod_{j=1}^c \left(\pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{(1-y_{ij})} \right) f(\eta_i) d\eta_i, \quad (\text{A.1.1})$$

$$\pi_i(\alpha, \Lambda) = \int \exp \left(\sum_{j=1}^c \left[y_{ij} \ln \pi_{ij} + (1 - y_{ij}) \ln (1 - \pi_{ij}) \right] - \frac{k}{2} \ln 2\pi - \frac{1}{2} \eta_i^T \eta_i \right) d\eta_i, \quad (\text{A.1.2})$$

Focusing on (A.1.2), we can obtain the marginal probability using Laplace approximation

$$\pi_i(\alpha, \Lambda) \approx \left[\prod_{j=1}^c \left(\exp(\alpha_j + \tilde{\eta}_i^T \Lambda_j) \right)^{y_{ij}} / \left(1 + \exp(\alpha_j + \tilde{\eta}_i^T \Lambda_j) \right) \right] |\Sigma_i|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \tilde{\eta}_i^T \tilde{\eta}_i \right), \quad i=1, 2, \dots, 2^c, \quad (\text{A.1.3})$$

$$\text{where } \Sigma_i = \left[-I^{(2)}(\tilde{\eta}_i) \right]^{-1} = \left(\sum_{j=1}^c \left[\pi_{ij} (1 - \pi_{ij}) \Lambda_j \Lambda_j^T \right] + I_k \right)^{-1}$$

A.2 The $\pi_i(\alpha, \Lambda)$ derivative with respect to parameters

$$\frac{\partial \pi_i(\alpha, \Lambda)}{\partial \alpha_j} = \pi_i(\alpha, \Lambda) \left\{ y_{ij} - \pi_{ij} - \frac{1}{2} \text{tr} \left(A_i^{-1} \frac{\partial A_i}{\partial \alpha_j} \right) \right\}, \quad j=1, 2, \dots, c, \quad (\text{A.2.1})$$

From the formula $A_i = \sum_{j=1}^c \left[\pi_{ij} (1 - \pi_{ij}) \eta_j \eta_j^T \right] + I_k$ we can obtained the first derivative about the α, Λ

$$\frac{\partial A_i}{\partial \lambda_{ij}} = \left[\pi_{ij} (1 - \pi_{ij}) - 2\pi_{ij}^2 (1 - \pi_{ij}) \right] \eta_{ij} \Lambda_j \Lambda_j^T + \left[\pi_{ij} (1 - \pi_{ij}) \right] \left(\frac{\partial \Lambda_j}{\partial \lambda_{ij}} \Lambda_j^T + \Lambda_j \frac{\partial \Lambda_j^T}{\partial \lambda_{ij}} \right), \quad t=1, 2, \dots, k, \quad j=1, 2, \dots, c, \quad (\text{A.2.2})$$

$$\frac{\partial A_i}{\partial \alpha_j} = \left[\pi_{ij} (1 - \pi_{ij}) - 2\pi_{ij}^2 (1 - \pi_{ij}) \right] \Lambda_j \Lambda_j^T, \quad j=1, 2, \dots, c, \quad (\text{A.2.3})$$

References

- Agresti, A. Categorical data analysis. 2nd edition. Wiley-Interscience; New York: 2002.
- An X, Bentler PM. Efficient direct sampling MCEM algorithm for latent variable models with binary responses. *Computational Statistics and Data Analysis*. 2012; 56(2):231–244.
- Baker, FB. Item response theory: Parameter estimation techniques. 2nd edition. Marcel Dekker; New York: 2004.
- Bartholomew, DJ.; Knott, M. Latent variable models and factor analysis. 2nd edition. Arnold ;Oxford University Press; London New York: 1999.
- Bartholomew DJ, Leung SO. A goodness of fit test for sparse 2^P contingency tables. *British Journal of Mathematical and Statistical Psychology*. 2002; 55:1–15. [PubMed: 12034008]
- Bishop, YMM.; Fienberg, SE.; Holland, PW. Discrete multivariate analysis: Theory and practice. MIT Press; Cambridge, Mass: 1975.
- Bjornstad JF. On the generalization of the likelihood function and the likelihood principle. *Journal of the American Statistical Association*. 1996; (434):791–806.
- Bock RD, Aitkin M. Marginal maximum-likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*. 1981; 46(4):443–459.

- Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993; 88(421):9–25.
- Butler RW. Predictive likelihood inference with applications. *Journal of the Royal Statistical Society Series B-Methodological*. 1986; 48(1):1–38.
- Cai L, Maydeu-Olivares A, Coffman DL, Thissen D. Limited-information goodness-of-fit testing of item response theory models for sparse 2^p tables. *British Journal of Mathematical & Statistical Psychology*. 2006; 59:173–194. [PubMed: 16709285]
- Chan JSK, Kuk AYC. Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics*. 1997; 53(1):86–97.
- Christofferson A. Factor analysis of dichotomized variables. *Psychometrika*. 1975; 40(1):5–32.
- Forero CG, Maydeu-Olivares A. Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*. 2009; 14(3):275–299. [PubMed: 19719362]
- Fu ZH, Tao J, Shi NZ, Zhang M, Lin N. Analyzing longitudinal item response data via the pairwise fitting method. *Multivariate Behavioral Research*. 2011; 46:669–690.
- Hoshino, T.; Bentler, PM. Bias in factor score regression and a simple solution. In: de Leon, AR.; Chough, KC., editors. *Analysis of mixed data*. Taylor & Francis; New York: 2012.
- Huber P, Ronchetti E, Victoria-Feser MP. Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*. 2004; 66:893–908.
- Joe H. Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*. 2008; 52:5066–5074.
- Joe H, Maydeu-Olivares A. A general family of limited information goodness-of-fit statistics for multinomial data. *Psychometrika*. 2010; 75(3):393–419.
- Khatri CG. A note on a Manova model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*. 1966; 18(1):75–86.
- Lee SY, Poon WY, Bentler PM. A two-stage estimation of structural equation models with continuous and polytomous variables. *British Journal of Mathematical and Statistical Psychology*. 1995; 48(2):339–358. [PubMed: 8527346]
- Lee Y, Nelder JA. Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series B-Methodological*. 1996; 58(4):619–656.
- Lee Y, Nelder JA. Double hierarchical generalized linear models. *Journal of the Royal Statistical Society Series C-Applied Statistics*. 2006; 55:139–167.
- Lee, Y.; Nelder, JA.; Pawitan, Y. *Generalized linear models with random effects: Unified analysis via h-likelihood*. Chapman and Hall/CRC; FL: 2006.
- Levine RA, Casella G. Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics*. 2001; 10(3):422–439.
- Lin XH, Breslow NE. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*. 1996; 91(435):1007–1016.
- Liu Q, Pierce DA. Heterogeneity in Mantel-Haenszel-type models. *Biometrika*. 1993; 80(3):543–556.
- Maydeu-Olivares A. Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika*. 2001; 66(2):209–227.
- Maydeu-Olivares A. Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*. 2006; 71(1):57–77.
- Maydeu-Olivares A, Joe H. Limited- and full-information estimation and goodness-of-fit testing in 2^n contingency tables: A unified framework. *Journal of the American Statistical Association*. 2005; 100(471):1009–1020.
- McDonald RP, Burr EJ. A comparison of four methods of constructing factor scores. *Psychometrika*. 1967; 32(4):381–401.
- Meng XL, Schilling S. Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*. 1996; 91(435):1254–1267.
- Meng XL, van Dyk D. The EM algorithm: An old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society Series B-Methodological*. 1997; 59(3):511–540.
- Meng XL, van Dyk D. Fast EM-type implementations for mixed effects models. *Journal of the Royal Statistical Society Series B-Statistical Methodology*. 1998; 60:559–578.

- Naylor JC, Smith AFM. Applications of a method for the efficient computation of posterior distributions. *Applied Statistics-Journal of the Royal Statistical Society Series C*. 1982; 31(3):214–225.
- Pinheiro JC, Chao EC. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*. 2006; 15(1): 58–81.
- Rao, CR. *Linear statistical inference and its applications*. 2nd edition. Wiley; New York: 1973.
- Raudenbush SW, Yang ML, Yosef M. Maximum likeli-hood for generalized linear models with nested random effects via highorder, multivariate laplace approximation. *Journal of Computational and Graphical Statistics*. 2000; 9(1):141–157.
- Reiser M. Analysis of residuals for the multinomial item response model. *Psychometrika*. 1996; 61(3): 509–528.
- Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye, IA.; Clogg, CC., editors. *Latent variables analysis: Applications for developmental research*. Vol. xii. Sage Publications, Inc; Thousand Oaks, Thousand Oaks, CA, US: 1994. p. 399-419.
- Schilling S, Bock RD. High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*. 2005; 70(3):533–555.
- Sheng YY. A MATLAB package for Markov Chain Monte Carlo with a multi-unidimensional IRT model. *Journal of Statistical Software*. 2008; 28(10):1–20.
- Skrondal A, Laake P. Regression among factor scores. *Psychometrika*. 2001; 66(4):563–575.
- Teugels JL. Some representations of the multivariate Bernoulli and binomial distributions. *Journal of Multivariate Analysis*. 1990; 32(2):256–268.
- Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*. 1986; 81(393):82–86.
- Wu, E.; Bentler, PM. *EQSIRT. Multivariate Software*; Encino, CA: 2011.
- Wu J, Bentler PM. Application of H-likelihood to factor analysis models with binary response data. *Journal of Multivariate Analysis*. 2012; 106(4):72–79.
- Yanai, H.; Ichikawa, M. Factor analysis. In: Rao, CR.; Sinharay, S., editors. *Handbook of statistics*. Vol. 26. Elsevier; Amsterdam: 2007. p. 257-296.
- Yuan KH, Bentler PM. Two simple approximations to the distributions of quadratic forms. *British Journal of Mathematical and Statistical Psychology*. 2010; 63(2):273–291. [PubMed: 19793410]
- Yun S, Lee Y. Comparison of hierarchical and marginal likeli-hood estimators for binary outcomes. *Computational Statistics and Data Analysis*. 2004; 45(3):639–650.

Table .1

Comparison of estimates of moment residuals up to 2nd and 3rd order

Parameter	True value	LIWLS 2		LIWLS 3	
		Estimate	SE	Estimate	SE
Intercepts	-0.4471	-0.4718	0.0354	-0.4675	0.0352
	0.2188	0.2359	0.0362	0.2214	0.0358
	-1.2565	-1.2083	0.0413	-1.1904	0.0413
	0.8201	0.8181	0.0446	0.7869	0.0435
	1.7741	1.7220	0.0545	1.7080	0.0566
	0.9569	0.9005	0.0384	0.8820	0.0389
	0.7803	0.8545	0.0413	0.8335	0.0410
	0.3831	0.3868	0.0353	0.3732	0.0353
	-1.6120	-1.6020	0.0541	-1.5668	0.0528
	0.1942	0.2570	0.0354	0.2413	0.0352
Coefficients	-1.0624	-1.0381	0.0409	-1.0200	0.0404
	0.3012	0.3091	0.0344	0.2937	0.0344
Intercepts RMSE		0.0401		0.0456	
Average of SE			0.0410		0.0409
Coefficients	0.6305	0.6056	0.0766	0.5909	0.0745
	0.7482	0.7753	0.0897	0.7450	0.0851
	0.5521	0.4664	0.0758	0.4798	0.0754
	0.8857	0.8678	0.1006	0.8544	0.0968
	0.8020	0.5847	0.0883	0.6596	0.0907
	0.5609	0.4857	0.0718	0.5352	0.0733
	0.7309	0.7169	0.0849	0.7167	0.0850
	0.6223	0.6231	0.0755	0.6295	0.0763
	0.5359	0.6771	0.0907	0.6623	0.0896
	0.7955	0.7061	0.0814	0.6894	0.0803
	0.6595	0.5624	0.0764	0.5473	0.0757
	0.6097	0.5710	0.0719	0.5812	0.0728

Parameter	True value	LIWLS 2		LIWLS 3	
		Estimate	SE	Estimate	SE
Coefficients RMSE		0.0917		0.0762	
Average of SE			0.0820		0.0813
Average RMSE		0.0708		0.0628	

Table .2

Comparison of estimates of the FIMDML, LIMDML 2 and LIMDML 3

Parameter	True value	LIMDML 2		LIMDML 3		FIMDML	
		Estimate	SE	Estimate	SE	Estimate	SE
Intercepts	-0.4471	-0.4707	0.0354	-0.4700	0.0353	-0.4687	0.0351
	0.2188	0.2348	0.0359	0.2341	0.0357	0.2348	0.0355
	-1.2565	-1.2189	0.0422	-1.2172	0.0420	-1.2205	0.0421
	0.8201	0.8081	0.0435	0.8074	0.0435	0.8094	0.0430
	1.7741	1.7392	0.0549	1.7491	0.0565	1.7979	0.0638
	0.9569	0.9080	0.0384	0.9073	0.0385	0.9081	0.0392
	0.7803	0.8584	0.0411	0.8579	0.0412	0.8440	0.0412
	0.3831	0.3885	0.0351	0.3874	0.0350	0.3721	0.0350
	-1.612	-1.6264	0.0569	-1.6295	0.0573	-1.6543	0.0586
	0.1942	0.2579	0.0351	0.2565	0.0350	0.2410	0.0349
Intercepts RMSE	-1.0624	-1.0394	0.0410	-1.0384	0.0409	-1.0517	0.0411
	0.3012	0.3114	0.0342	0.3104	0.0342	0.2956	0.0342
Average of SE		0.0376	0.0411	0.0370	0.0413	0.0334	0.0420
Coefficients	0.6305	0.5994	0.0776	0.5919	0.0773	0.584	0.0742
	0.7482	0.7416	0.0889	0.7268	0.0876	0.7075	0.0822
	0.5521	0.4990	0.0785	0.4914	0.0783	0.511	0.0773
	0.8857	0.8137	0.0979	0.8100	0.0978	0.7998	0.0922
	0.802	0.5800	0.0900	0.6142	0.0920	0.7724	0.0994
	0.5609	0.4849	0.0730	0.4864	0.0732	0.5352	0.0734
	0.7309	0.7012	0.0849	0.7044	0.0855	0.7189	0.0855
	0.6223	0.6042	0.0751	0.6000	0.0751	0.6072	0.0746
	0.5359	0.7267	0.0953	0.7322	0.0960	0.7558	0.0969
	0.7955	0.6775	0.0802	0.6682	0.0798	0.6617	0.0782
Coefficients	0.6595	0.5623	0.0770	0.5553	0.0768	0.5602	0.0763
	0.6097	0.5511	0.0714	0.5486	0.0715	0.5593	0.0713

Parameter	True value	LIMDML 2		LIMDML 3		FIMDML	
		Estimate	SE	Estimate	SE	Estimate	SE
Coefficients RMSE		0.1035		0.1011		0.0822	
Average of SE			0.0825		0.0826		0.0818
Average RMSE		0.0779		0.0761		0.0667	

Table .3

Comparison of estimates of MCMC, MML, and LAML

Parameter	True value	MCMC	MML	LAML
		Estimate	Estimate	Estimate
Intercepts	-0.4471	-0.5156	-0.4855	-0.4687
	0.2188	0.2333	0.2256	0.2348
	-1.2565	-1.2562	-1.2214	-1.2205
	0.8201	0.8469	0.8188	0.8093
	1.7741	1.7214	1.7264	1.7980
	0.9569	0.9208	0.8919	0.9081
	0.7803	0.8767	0.8517	0.8440
	0.3831	0.3876	0.3763	0.3722
	-1.6120	-1.6304	-1.6190	-1.6538
	0.1942	0.2469	0.2467	0.2410
	-1.0624	-1.0813	-1.0506	-1.0517
	0.3012	0.3099	0.2985	0.2957
Interceptsts RMSE		0.0434	0.0380	0.0334
Coefficients	0.6305	0.6039	0.5870	0.5839
	0.7482	0.8027	0.7744	0.7074
	0.5521	0.4573	0.4536	0.5112
	0.8857	0.8673	0.8389	0.7996
	0.8020	0.5620	0.6158	0.7727
	0.5609	0.5045	0.5052	0.5353
	0.7309	0.7184	0.7085	0.7188
	0.6223	0.6326	0.6209	0.6072
	0.5359	0.6117	0.6360	0.7549
	0.7955	0.7177	0.7005	0.6618
	0.6595	0.5177	0.5400	0.5604
	0.6097	0.5642	0.5744	0.5594
Coefficients RMSE		0.0949	0.0853	0.0880
Average RMSE		0.0738	0.0660	0.0666

Table 4

Comparison of test statistics

	df	M2	p2	df	M3	p3	df	χ^2	P
LIWLS 2	54	59.71	0.2760	274	277.26	0.4335	4071	3743.00	0.9999
LIWLS 3	54	58.15	0.3251	274	270.27	0.5523	4071	3585.90	1.0000
FIMDML	54	61.06	0.2372	274	280.03	0.3881	4071	3798.56	0.9990
LIMDML 2	54	60.55	0.2515	274	280.11	0.3869	4071	3785.14	0.9994
LIMDML 3	54	60.65	0.2486	274	280.28	0.3841	4071	3781.36	0.9995
MCMC	54	60.67	0.2479	274	284.38	0.3206	4071	3945.13	0.9196
MML	54	60.16	0.2627	274	279.56	0.3957	4071	3769.88	0.9997
LAML	54	61.06	0.2373	274	280.03	0.3882	4071	3798.4f	0.9990

Table .5

Simulation 1: Performance of two limited information estimators

Parameter	True value	LIWLS 2	LIWLS 3
		Estimate	Estimate
Intercepts	-0.4471	0.0020	0.0022
	0.2188	0.0023	0.0030
	-1.2565	0.0031	0.0041
	0.8201	0.0040	0.0043
	1.7741	0.0105	0.0080
	0.9569	0.0027	0.0030
	0.7803	0.0026	0.0029
	0.3831	0.0021	0.0028
	-1.6120	0.0050	0.0057
	0.1942	0.0022	0.0024
	-1.0624	0.0044	0.0025
	0.3012	0.0026	0.0027
Average of MSSE		0.0036	0.0036
Coefficients	0.6305	0.0089	0.0086
	0.7482	0.0104	0.0077
	0.5521	0.0126	0.0095
	0.8857	0.0167	0.0116
	0.8020	0.0229	0.0189
	0.5609	0.0085	0.0080
	0.7309	0.0130	0.0079
	0.6223	0.0090	0.0070
	0.5359	0.0151	0.0114
	0.7955	0.0100	0.0053
	0.6595	0.0081	0.0090
	0.6097	0.0066	0.0042
Average of MSSE		0.0118	0.0091

Table .6

Simulation 2: Null performance of three test statistics

significance level	0.1						0.05					
Test-Statistics	χ^2		$L2$		$L3$		χ^2		$L2$		$L3$	
Proportion less than significance level	0.382		0.179		0.180		0.350		0.111		0.108	