Comparison between conditional and marginal maximum likelihood for a class of item response models

Francesco Bartolucci, University of Perugia (IT)
Silvia Bacci, University of Perugia (IT)
Claudia Pigini, University of Perugia (IT)

ASMOD 2013
Napoli - November, 25-26, 2013
Outline

Motivation and purpose
Graded Response Model
Standard methods for maximum likelihood estimation
Proposed conditional maximum likelihood method
Simulation study of the proposed estimator
Hausman test for normality of the latent trait
Application
Conclusions
References
Motivation and purpose

- In the literature on *latent variable models*, there is a considerable interest in estimation methods that do not require parametric assumptions on the latent distribution.

- We focus on an Item Response Theory model for *ordinal responses* which is known as Graded Response Model.

- We introduce a *conditional likelihood estimator* which requires no assumptions on the latent distribution and is very simple to implement.

- The method also allows us to implement a *Hausman test* for a parametric assumption (e.g., normal distribution) on the latent distribution.
Graded Response Model (GRM)

- For a questionnaire of \( r \) items, let \( X_j \) denote the response variable for the \( j \)-th item \((j = 1, \ldots, r)\), which is assumed to have \( l_j \) categories, indexed from 0 to \( l_j - 1 \).

- **Assumptions** of the GRM model (Samejima, 1969):
  - **unidimensionality**: the test items contribute to measure a single latent trait \( \Theta \) corresponding to a type of ability in education.
  - **local independence**: the response variables \( X_1, \ldots, X_r \) are conditionally independent given \( \Theta \):
    \[
    p(x_1, \ldots, x_r | \theta) = \prod_{j=1}^{r} p(x_j | \theta)
    \]
  - **monotonicity**: \( p(X_j \geq x | \theta) \) is nondecreasing in \( \theta \) for all \( j \):
    \[
    \log \frac{p(X_j \geq x | \theta)}{p(X_j < x | \theta)} = \gamma_j (\theta - \beta_{jx}), \quad x = 1, \ldots, l_j - 1
    \]
- $\gamma_j$ identifies the *discriminating power* of item $j$ (typically $\gamma_j > 0$)

- $\beta_{jx}$ denotes the *difficulty level* for item $j$ and category $x$, ordered as $\beta_{j1} < \ldots < \beta_{j,l-1}$

- We focus on a special case of GRM (1P-GRM) in which *all the items discriminate* in the same way (van der Ark, 2001):

  $$\gamma_1 = \cdots = \gamma_r = 1$$

- We also consider a further special case (1P-RS-GRM) based on the *rating scale parametrization* (items have the same number of response categories):

  $$\beta_{jx} = \beta_j + \tau_x, \quad j = 1, \ldots, r, \ x = 1, \ldots, l - 1,$$

  where $\beta_j$ represents the difficulty of item $j$ and $\tau_x$ are cut-points common to all items
Maximum likelihood estimation

- Given a sample of observations $x_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, r$, different *maximum likelihood estimation methods* may be used.

- Under a fixed-effects formulation, the model may be estimated by the **Joint Maximum Likelihood** (JML) method based on:

$$
\ell_J(\lambda) = \sum_{i=1}^{n} \log \prod_{j=1}^{r} p(x_{ij}|\theta_i) = \sum_{i=1}^{n} \sum_{j=1}^{r} \log p(x_{ij}|\theta_i)
$$

with the parameter vector $\lambda$ also including the ability parameters $\theta_i$.

- The JML method is simple to implement but *it does not ensure consistency* of the parameter estimates and may suffer from instability problems.
Under a random-effects formulation, with the latent trait assumed to have a normal distribution, we can use the Marginal Maximum Likelihood (MML) method based on:

\[ \ell_M(\eta) = \sum_{i=1}^{n} \log \int \phi(\theta_i; 0, \sigma^2) \prod_{j=1}^{r} p(x_{ij}|\theta_i) d\theta_i \]

with \( \phi(\theta_i; 0, \sigma^2) \) denoting the density function of \( N(0, \sigma^2) \) and the parameter vector \( \eta \) containing the item parameters and \( \sigma^2 \).

The MML method is more complex to implement (requires a quadrature for the integral) and the parameter estimates are consistent under the hypothesis of normality of the latent trait.

In order to reduce the dependence of the parameter estimates on parametric assumptions on the latent distribution, we can use a semi-parametric method (MML-LC) based on the assumption that the latent trait has a discrete distribution with \( k \) support points (latent classes).
The MML-LC method is based on the *marginal log-likelihood function*:

\[
\ell_{LC}(\psi) = \sum_{i=1}^{n} \log \sum_{c=1}^{k} \pi_c \prod_{j=1}^{r} p(x_{ij} | \theta_i = \xi_c)
\]

with \(\xi_1, \ldots, \xi_k\) being the support points and \(\pi_1, \ldots, \pi_k\) the corresponding mass probabilities; these are included in the parameter vector \(\psi\) together with the item parameters.

The *EM algorithm* (Dempster et al., 1977) is typically used for the maximization of \(\ell_{LC}(\psi)\).

A drawback of the method is the greater *numerical complexity* and the need to *choose \(k\) properly* (AIC and BIC may be used in this regard).

Some *instability problems* may arise with large values of \(k\).
Conditional maximum likelihood method

We suggest a *Conditional Maximum Likelihood* (CML) method based on considering all the possible dichotomizations of the response variables (Baetschmann et al., 2011)

For the case in which the response variables have the *same number* *l* of response categories:

1. we consider the *l* − 1 dichotomizations indexed by *d* = 1, . . . , *l* − 1
2. for each dichotomization *d* we transform the response variables *X*ₖ (for every unit) in the binary variables

   \[ Y_{j}^{(d)} = 1\{X_{j} \geq d\}, \quad j = 1, \ldots, r, \]

   with \(1\{\cdot\}\) being the indicator function

3. we maximize the function given by the *sum of the conditional log-likelihood functions* (Anderson, 1973) corresponding to each dichotomization:

\[
\ell_{C}^{*}(\beta) = \sum_{d=1}^{l-1} \log p(y_{i1}^{(d)}, \ldots, y_{ir}^{(d)} | y_{i+}^{(d)}), \quad y_{i+}^{(d)} = \sum_{j=1}^{r} y_{ij}^{(d)}
\]
The method relies on the fact that the dichotomized variable distributions satisfy the Rasch (1961) model:

\[
\log \frac{p(Y_j^{(d)} = 1 | \theta)}{p(Y_j^{(d)} = 0 | \theta)} = \theta - \beta_{jd}, \quad j = 1, \ldots, r, \quad d = 1, \ldots, l - 1
\]

The total score \( Y_+^{(d)} = \sum_{j=1}^{r} Y_j^{(d)} \) is a sufficient statistic for the ability parameter \( \theta \).

The resulting conditional probability involved in \( \ell^*_C(\beta) \) has expression:

\[
p(y_{i1}^{(d)}, \ldots, y_{ir}^{(d)} | y_{i+}^{(d)}) = \frac{\exp \left( - \sum_{j=1}^{r} y_{ij}^{(d)} \beta_{jx} \right)}{\sum_{\text{z: } z_+ = y_{i+}^{(d)}} \exp \left( - \sum_{j=1}^{r} z_j \beta_{jx} \right)}
\]

with \( \sum_{\text{z: } z_+ = y_{i+}^{(d)}} \) extended to all binary vectors \( z \) of dimension \( r \) with elements summing up to \( y_{i+}^{(d)} \).
The likelihood function \( \ell^*_C(\beta) \) depends only on the item parameters (\( \beta_{jx} \) or \( \beta_j \)) collected in \( \beta \):

- under 1P-GRM the identifiable parameters are \( \beta_{jx} \) for \( j = 2, \ldots, r \) and \( x = 1, \ldots, l - 1 \) (we use the constraint \( \beta_{1x} = 0, x = 1, \ldots, l - 1 \))

- under 1P-RS-GRM the identifiable parameters are \( \beta_j \) for \( j = 2, \ldots, r \) (we use the constraint \( \beta_1 = 0 \)), whereas the cut-points \( \tau_x \) are not identified

This function may be simply maximized by a Newton-Raphson algorithm based on:

- pseudo score vector:
  \[
  s^*_C(\beta) = \sum_{i=1}^{n} s^*_{C,i}(\beta), \quad s^*_{C,i}(\beta) = \frac{\partial}{\partial \beta} \log p(y_{i1}^{(d)}, \ldots, y_{ir}^{(d)} | y_{i+}^{(d)})
  \]

- pseudo observed information matrix:
  \[
  H^*_C(\beta) = -\sum_{i=1}^{n} \frac{\partial^2}{\partial \beta \partial \beta'} \log p(y_{i1}^{(d)}, \ldots, y_{ir}^{(d)} | y_{i+}^{(d)})
  \]
The **asymptotic variance-covariance matrix** may be obtained by the sandwich formula:

\[
\hat{V}_C^*(\hat{\beta}_C^*) = H_C^*(\hat{\beta}_C^*)^{-1}S(\hat{\beta}_C^*)H_C^*(\hat{\beta}_C^*)^{-1}
\]

\[
S(\beta) = \sum_{i=1}^{n} s^*_C,i(\beta)[s^*_C,i(\beta)]'
\]

**Standard errors** may be extracted in the usual way from \( \hat{V}_C^*(\hat{\beta}_C^*) \)

On the basis of the pseudo score vector and information we can also implement a *Hausman (1978) test* for the hypothesis of normality in which the estimate \( \hat{\beta}_C^* \) is compared with the corresponding estimate obtained from the MML method.
Simulation study of the CML estimator

- We *simulated* 1,000 samples of size $n$ from the 1P-RS-GRM model for $r$ response variable with $l = 5$ categories:
  - $r = 5, 10$
  - $n = 1000, 2000$
  - cut-points ($\tau_x$) equal to $-2, -0.5, 0.5, 2$
  - difficulty parameters ($\beta_j$) as $r$ equally distant points in $[-2, 2]$
  - four different latent distributions (all are standardized):
    - Normal(0,1)
    - Gamma(2,2)
    - LC1: latent class model with symmetric distribution based on mass probabilities 0.25, 0.5, 0.25 for increasing and equally spaced support points
    - LC2: as in LC1 but with skewed distribution based on mass probabilities 0.4, 0.5, 0.1

- For *all samples* we fit 1P-GRM and 1P-RS-GRM by the MML, MML-LC ($k$ chosen by BIC), and CML methods
Simulation results for 1P-GRM: average values of absolute bias and RMSE for the estimates of parameters $\beta_{jx}$

<table>
<thead>
<tr>
<th>Distrib.</th>
<th>$n$</th>
<th>$r$</th>
<th>CML abs.bias</th>
<th>CML RMSE</th>
<th>MML abs.bias</th>
<th>MML RMSE</th>
<th>MML-LC abs.bias</th>
<th>MML-LC RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(0, 1)$</td>
<td>1000</td>
<td>5</td>
<td>0.0121</td>
<td>0.1646</td>
<td>0.0112</td>
<td>0.1575</td>
<td>0.0019</td>
<td>0.1569</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0043</td>
<td>0.1134</td>
<td>0.0032</td>
<td>0.1080</td>
<td>0.0089</td>
<td>0.1081</td>
</tr>
<tr>
<td>$N(0, 1)$</td>
<td>1000</td>
<td>10</td>
<td>0.0085</td>
<td>0.1549</td>
<td>0.0085</td>
<td>0.1521</td>
<td>0.0156</td>
<td>0.1514</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0041</td>
<td>0.1086</td>
<td>0.0038</td>
<td>0.1069</td>
<td>0.0216</td>
<td>0.1083</td>
</tr>
<tr>
<td>$\Gamma(2, 2)$</td>
<td>1000</td>
<td>5</td>
<td>0.0070</td>
<td>0.1640</td>
<td>0.0634</td>
<td>0.1721</td>
<td>0.0053</td>
<td>0.1568</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0025</td>
<td>0.1139</td>
<td>0.0618</td>
<td>0.1306</td>
<td>0.0080</td>
<td>0.1098</td>
</tr>
<tr>
<td>$\Gamma(2, 2)$</td>
<td>1000</td>
<td>10</td>
<td>0.0150</td>
<td>0.1573</td>
<td>0.0474</td>
<td>0.1639</td>
<td>0.0128</td>
<td>0.1543</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0087</td>
<td>0.1088</td>
<td>0.0455</td>
<td>0.1189</td>
<td>0.0138</td>
<td>0.1074</td>
</tr>
<tr>
<td>LC1</td>
<td>1000</td>
<td>5</td>
<td>0.0109</td>
<td>0.1619</td>
<td>0.0221</td>
<td>0.1586</td>
<td>0.0071</td>
<td>0.1572</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0068</td>
<td>0.1126</td>
<td>0.0183</td>
<td>0.1101</td>
<td>0.0059</td>
<td>0.1077</td>
</tr>
<tr>
<td>LC1</td>
<td>1000</td>
<td>10</td>
<td>0.0056</td>
<td>0.1553</td>
<td>0.0144</td>
<td>0.1545</td>
<td>0.0059</td>
<td>0.1526</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0031</td>
<td>0.1068</td>
<td>0.0099</td>
<td>0.1063</td>
<td>0.0031</td>
<td>0.1050</td>
</tr>
<tr>
<td>LC2</td>
<td>1000</td>
<td>5</td>
<td>0.0115</td>
<td>0.1650</td>
<td>0.0305</td>
<td>0.1634</td>
<td>0.0080</td>
<td>0.1587</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0044</td>
<td>0.1157</td>
<td>0.0251</td>
<td>0.1163</td>
<td>0.0039</td>
<td>0.1116</td>
</tr>
<tr>
<td>LC2</td>
<td>1000</td>
<td>10</td>
<td>0.0089</td>
<td>0.1569</td>
<td>0.0199</td>
<td>0.1573</td>
<td>0.0084</td>
<td>0.1544</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0033</td>
<td>0.1104</td>
<td>0.0174</td>
<td>0.1117</td>
<td>0.0034</td>
<td>0.1089</td>
</tr>
</tbody>
</table>
Simulation results for 1P-RS-GRM: average values of absolute bias and RMSE for the estimates of parameters $\beta_j$

<table>
<thead>
<tr>
<th>Distrib.</th>
<th>$n$</th>
<th>$r$</th>
<th>CML</th>
<th></th>
<th>MML</th>
<th></th>
<th>MML-LC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>abs.bias</td>
<td>RMSE</td>
<td>abs.bias</td>
<td>RMSE</td>
<td>abs.bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>$N(0, 1)$</td>
<td>1000</td>
<td>5</td>
<td>0.0042</td>
<td>0.1005</td>
<td>0.0007</td>
<td>0.0955</td>
<td>0.0055</td>
<td>0.0960</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0012</td>
<td>0.0693</td>
<td>0.0030</td>
<td>0.0645</td>
<td>0.0078</td>
<td>0.0653</td>
</tr>
<tr>
<td>$N(0, 1)$</td>
<td>1000</td>
<td>10</td>
<td>0.0022</td>
<td>0.0923</td>
<td>0.0040</td>
<td>0.0936</td>
<td>0.0168</td>
<td>0.0902</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0013</td>
<td>0.0637</td>
<td>0.0030</td>
<td>0.0603</td>
<td>0.0199</td>
<td>0.0647</td>
</tr>
<tr>
<td>$\Gamma(2, 2)$</td>
<td>1000</td>
<td>5</td>
<td>0.0000</td>
<td>0.0988</td>
<td>0.0130</td>
<td>0.0945</td>
<td>0.0075</td>
<td>0.0940</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0015</td>
<td>0.0690</td>
<td>0.0125</td>
<td>0.0648</td>
<td>0.0105</td>
<td>0.0663</td>
</tr>
<tr>
<td>$\Gamma(2, 2)$</td>
<td>1000</td>
<td>10</td>
<td>0.0078</td>
<td>0.0920</td>
<td>0.0072</td>
<td>0.0861</td>
<td>0.0109</td>
<td>0.0890</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0046</td>
<td>0.0648</td>
<td>0.0108</td>
<td>0.0644</td>
<td>0.0154</td>
<td>0.0640</td>
</tr>
<tr>
<td>LC1</td>
<td>1000</td>
<td>5</td>
<td>0.0000</td>
<td>0.0978</td>
<td>0.0043</td>
<td>0.0905</td>
<td>0.0020</td>
<td>0.0945</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0037</td>
<td>0.0693</td>
<td>0.0040</td>
<td>0.0640</td>
<td>0.0025</td>
<td>0.0650</td>
</tr>
<tr>
<td>LC1</td>
<td>1000</td>
<td>10</td>
<td>0.0021</td>
<td>0.0947</td>
<td>0.0069</td>
<td>0.0968</td>
<td>0.0019</td>
<td>0.0801</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0011</td>
<td>0.0647</td>
<td>0.0036</td>
<td>0.0647</td>
<td>0.0012</td>
<td>0.0620</td>
</tr>
<tr>
<td>LC2</td>
<td>1000</td>
<td>5</td>
<td>0.0040</td>
<td>0.1003</td>
<td>0.0095</td>
<td>0.0955</td>
<td>0.0008</td>
<td>0.0953</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>5</td>
<td>0.0028</td>
<td>0.0718</td>
<td>0.0082</td>
<td>0.0705</td>
<td>0.0038</td>
<td>0.0678</td>
</tr>
<tr>
<td>LC2</td>
<td>1000</td>
<td>10</td>
<td>0.0038</td>
<td>0.0951</td>
<td>0.0063</td>
<td>0.0844</td>
<td>0.0032</td>
<td>0.0819</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>10</td>
<td>0.0007</td>
<td>0.0662</td>
<td>0.0044</td>
<td>0.0608</td>
<td>0.0011</td>
<td>0.0638</td>
</tr>
</tbody>
</table>
Main conclusions from the simulation study

- **Very similar performances** are observed in terms of efficiency under the normal distribution (the MML method is the most efficient, but the RMSE of the CML estimator is rather close).

- **A certain bias** arises for the MML method when the distribution is not normal (especially in the Gamma(2,2) case), whereas this bias is negligible for the CML method and the MML-LC method.

- When the latent distribution is not normal, and then the MML estimator is biased, the CML method performs very similarly to the MML-LC method, with a **negligible loss of efficiency** of the CML method.
Hausman test for normality of the latent trait

- The hypothesis of normality on which the MML method is based may be tested by a *Hausman test statistic*:

\[
T = (\hat{\beta}_M - \hat{\beta}_C)' \hat{\mathbf{W}}^{-1} (\hat{\beta}_M - \hat{\beta}_C)
\]

with \( \hat{\beta}_M \) being the estimator based on the MML method under the constraint \( \beta_{1x} = 0, \quad x = 1, \ldots, l - 1 \)

- \( \hat{\mathbf{W}} \) is the *estimate of the variance-covariance matrix* of \( \hat{\beta}_M - \hat{\beta}_C \) obtained starting from the sandwich formula (\( \hat{\beta}_M \) is a function of \( \hat{\lambda}_M \)):

\[
\hat{\mathbf{V}}(\hat{\lambda}_M, \hat{\beta}_C) = \left( \begin{array}{cc} H_M(\hat{\lambda}_M) & O \\ O & H^*_C(\hat{\beta}_C) \end{array} \right)^{-1} S^*(\hat{\lambda}_M) \left( \begin{array}{cc} H_M(\hat{\lambda}_M) & O \\ O & H^*_C(\hat{\beta}_C) \end{array} \right)^{-1}
\]

\[
S^*(\hat{\lambda}_M, \hat{\beta}_C) = \sum_{i=1}^n \begin{pmatrix} s_{M,i}(\hat{\lambda}_M) \\ s_{C,i}(\hat{\beta}_C) \end{pmatrix} \begin{pmatrix} s_{M,i}(\hat{\lambda}_M)' & s_{C,i}(\hat{\beta}_C)' \end{pmatrix}
\]
Under the 1P-GRM model, the asymptotic null distribution of $T$ is $\chi^2((r - 1)(l - 1))$.

Under the 1P-RS-GRM model, the asymptotic null distribution of $T$ is $\chi^2(r - 1)$.

If the hypothesis of normality is rejected, we estimate the model in a semi-parametric way by the MML-LC method.
Application

- We consider a dataset (available in R) referred to a sample of $n = 392$ individuals from UK extracted from the Consumer Protection and Perceptions of Science and Technology section of the 1992 Euro-Barometer Survey.

- The dataset is based on the responses to $r = 7$ items (with $l = 4$ ordered categories):
  - **Comfort**: Science and technology are making our lives healthier, easier and more comfortable.
  - **Environment**: Scientific and technological research cannot play an important role in protecting the environment and repairing it.
  - **Work**: The application of science and new technology will make work more interesting.
  - **Future**: Thanks to science and technology, there will be more opportunities for the future generations.
  - **Technology**: New technology does not depend on basic scientific research.
  - **Industry**: Scientific and technological research do not play an important role in industrial development.
  - **Benefit**: The benefits of science are greater than any harmful effect it may have.
**Estimation results of CML and MML methods** (under the constraint $\beta_{1x} = 0, \ x = 1, \ldots, l - 1$)

<table>
<thead>
<tr>
<th></th>
<th>1st cut-point</th>
<th>2nd cut-point</th>
<th>3rd cut-point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CML</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td>1.966 (.487)</td>
<td>1.531 (.211)</td>
<td>-0.628 (.189)</td>
</tr>
<tr>
<td>Work</td>
<td>2.125 (.468)</td>
<td>1.688 (.208)</td>
<td>0.698 (.197)</td>
</tr>
<tr>
<td>Future</td>
<td>1.115 (.488)</td>
<td>1.051 (.198)</td>
<td>-0.121 (.183)</td>
</tr>
<tr>
<td>Technology</td>
<td>1.401 (.529)</td>
<td>1.395 (.202)</td>
<td>-0.598 (.195)</td>
</tr>
<tr>
<td>Industry</td>
<td>0.742 (.577)</td>
<td>0.514 (.220)</td>
<td>-1.121 (.189)</td>
</tr>
<tr>
<td>Benefit</td>
<td>1.580 (.425)</td>
<td>1.558 (.200)</td>
<td>0.203 (.185)</td>
</tr>
<tr>
<td>Log-lik.</td>
<td>-1734.413</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MML</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td>1.885 (.486)</td>
<td>1.533 (.215)</td>
<td>-0.609 (.170)</td>
</tr>
<tr>
<td>Work</td>
<td>2.049 (.465)</td>
<td>1.716 (.213)</td>
<td>0.623 (.183)</td>
</tr>
<tr>
<td>Future</td>
<td>1.086 (.479)</td>
<td>1.076 (.203)</td>
<td>-0.116 (.168)</td>
</tr>
<tr>
<td>Technology</td>
<td>1.357 (.524)</td>
<td>1.394 (.207)</td>
<td>-0.576 (.176)</td>
</tr>
<tr>
<td>Industry</td>
<td>0.719 (.563)</td>
<td>0.499 (.227)</td>
<td>-1.013 (.167)</td>
</tr>
<tr>
<td>Benefit</td>
<td>1.524 (.424)</td>
<td>1.590 (.207)</td>
<td>0.169 (.171)</td>
</tr>
<tr>
<td>Log-lik.</td>
<td>-3014.706</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Hausman test leads to *reject the hypothesis of normality*:

\[ T = 39.9106, \quad \text{Prob}(\chi^2_{18} > T) = 0.002146 \]

We then estimate the model by the *MML-LC method with k = 3 latent classes* obtaining:

<table>
<thead>
<tr>
<th>( c )</th>
<th>( \hat{\xi}_c )</th>
<th>( \hat{\pi}_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.158</td>
<td>0.265</td>
</tr>
<tr>
<td>2</td>
<td>-0.073</td>
<td>0.548</td>
</tr>
<tr>
<td>3</td>
<td>1.851</td>
<td>0.187</td>
</tr>
</tbody>
</table>

The latent distribution is standardized and *skewed* (skewness index = 0.777)
Estimation results from the MML-LC method with $k = 3$

<table>
<thead>
<tr>
<th></th>
<th>1st cut-point</th>
<th>2nd cut-point</th>
<th>3rd cut-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>1.848 (.537)</td>
<td>1.497 (.282)</td>
<td>-0.623 (.182)</td>
</tr>
<tr>
<td>Work</td>
<td>2.011 (.528)</td>
<td>1.682 (.293)</td>
<td>0.639 (.185)</td>
</tr>
<tr>
<td>Future</td>
<td>1.067 (.480)</td>
<td>1.050 (.225)</td>
<td>-0.116 (.164)</td>
</tr>
<tr>
<td>Technology</td>
<td>1.332 (.519)</td>
<td>1.371 (.262)</td>
<td>-0.582 (.212)</td>
</tr>
<tr>
<td>Industry</td>
<td>0.701 (.602)</td>
<td>0.493 (.203)</td>
<td>-1.030 (.219)</td>
</tr>
<tr>
<td>Benefit</td>
<td>1.506 (.479)</td>
<td>1.557 (.282)</td>
<td>0.174 (.158)</td>
</tr>
<tr>
<td>Log-lik.</td>
<td></td>
<td>-3010.826</td>
<td></td>
</tr>
</tbody>
</table>

- The estimates of the item parameters are rather similar with respect to the MML method and the log-likelihood is higher.

- The influence on prediction of the latent ability may be considerable (prediction for a certain subject on the basis of the sequence of responses he/she provided through a posterior expected value).
Conclusions

- The proposed method for estimating the parameters of a constrained version of GRM is very simple to implement and is consistent under any true distribution of the latent trait.

- The method seems to provide an efficient estimator (efficiency close to the MML estimator under the normal distribution).

- It also allows us to implement a Hausman test for the hypothesis of normality.

- When the hypothesis of normality is rejected, the semi-parametric MML-LC method is an interesting alternative to MML.

- Even if significant differences are not observed in terms of estimates of the item parameters, the effect on prediction of the ability levels may be relevant.
References