Measurement and Research Department Reports 2000-4

EQUIVALENT LINEAR LOGISTIC TEST MODELS

T.M. Bechger
N.D. Verhelst
H.H.F.M. Verstralen

Cito

Measurement and Research Department Reports 2000-4

EQUIVALENT LINEAR LOGISTIC TEST MODELS

T.M. Bechger
H.H.F.M. Verstralen
N.D. Verhelst

Cito
Arnhem, 2000

Abstract

This paper is about the Linear Logistic Test Model (LLTM). We demonstrate that there are infinitely many equivalent ways to specify a model. An implication is that there may well be many ways to change the specification of a given LLTM and achieve the same improvement in model fit. To illustrate this phenomenon we analyze a real data set using a Lagrange multiplier test for the specification of the model.

1 Introduction

Let X_{vi} denote a dichotomous random variable with realization x_{vi} that indicates that subject v chooses a particular response to item i. For ease of presentation we will assume that X_{vi} is coded as follows:

$$x_{vi} = \begin{cases} 1 \text{ if the response is correct} \\ 0 \text{ if the response is wrong} \end{cases} . \tag{1}$$

In *the Rasch Model* (RM), the probability of a correct response is defined as

$$\Pr(X_{vi} = x_{vi}|\xi_v, \beta_i) = \frac{\exp[x_{vi}(\xi_v - \beta_i)]}{1 + \exp(\xi_v - \beta_i)}, \tag{2}$$

where $-\infty < \xi_v, \beta_i < \infty$. The probability of a correct answer is decreasing in β_i, which is recognized to be an item difficulty parameter. Similarly, ξ_v is a person parameter.

The parameters of the RM are not identified because $\Pr(X_{vi} = x_{vi}|\xi_v + c, \beta_i + c) = \Pr(X_{vi} = x_{vi}|\xi_v, \beta_i)$ for any constant c, and for all $i \in \{1, ..., k\}$. It is customary to remove this arbitrariness in the parameterization of the model by imposing a linear restriction on the item parameters. This restriction is customary called a normalization.

Definition 1 *A normalization is a linear restriction on the item parameters, i.e.,*

$$\sum_{i=1}^{k} a_i\beta_i = d, \tag{3}$$

where $\sum_{i=1}^{k} a_i = 1$, *and* d *is an arbitrary constant that is given the value zero.*

This restriction implies that there are only $k-1$ independent item parameters, and the item parameters must be interpreted relative to $\sum_{i=1}^{k} a_i\beta_i$. It is often convenient, for example, to consider the value of one of the item parameters as a reference. This amounts to setting $a_g = 1$ (for some $g \in \{1, ..., k\}$) and $a_i = 0$ $(i \neq g)$.

Let $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^t$, where uppercase t denotes transposition. The LLTM is a RM with linear restrictions on the normalized item parameters, that is,

$$\boldsymbol{\beta} = \mathbf{Q}\boldsymbol{\eta} = (\sum_{l=1}^{p} q_{il}\eta_l), \tag{4}$$

where $\mathbf{Q} = (q_{is} : i \in \{1, ..., k\}, s \in \{1, ..., p\})$ is a $k \times p$ matrix of constant weights, and $\boldsymbol{\eta} = (\eta_s : s \in \{1, ..., p\})$ is a p-dimensional vector of so-called basic parameters (see Fischer, 1995, and references therein). We assume that $p < (k - 1)$ so that the model is more parsimonious than the RM. We also assume that there are no restrictions on $\boldsymbol{\eta}$.

In many applications, the LLTM is specified according to a formalized theory about the cognitive operations required to solve the items, and the basic parameters represent the difficulty of particular cognitive operations. Consider, as an illustration, a set of paragraph comprehension items, which are classified according to the "transformations" the students use in responding. For example, with some questions the answer resides in the text, in much the same wording as in the question, while in others, subjects would have to go beyond the information given and make inferences in arriving at the correct answer. Under the assumptions implied by the RM, the difficulty of each transformation can be represented by a basic parameter and the (g, l) entry in the $\mathbf{Q}$-matrix represents the number of times the lth transformation is required to answer the gth item. Provided the theory is appropriate, and the $\mathbf{Q}$-matrix is specified correctly, the LLTM furnishes a useful extension of the RM. Especially, because the model enables one to predict the difficulty of new items which may facilitate the construction of new tests for different groups of students. When the $\mathbf{Q}$-matrix is misspecified, however, the predictions based on the model may be erroneous.

This example illustrates the common practice to specify the $\mathbf{Q}$-matrix for k item parameters, while the need to impose a normalization is ignored. Fischer (e.g., 1995) suggests that we examine the $\mathbf{Q}$ -matrix to assure ourselves that a normalization is imposed. To be more precise, he argues that the LLTM should be formulated as

$$\boldsymbol{\beta} = \mathbf{Q}\boldsymbol{\eta} + c\mathbf{1}_k = \begin{pmatrix} \mathbf{Q} & \mathbf{1}_k \end{pmatrix} \begin{pmatrix} \eta \\ c \end{pmatrix}, \tag{5}$$

where $\mathbf{1}_k$ represents a k-dimensional unit vector, and the constant c is added to indicate that the item parameters are not unique. A normalization implies that c is uniquely determined, which is true if and only if the augmented matrix $\begin{pmatrix} \mathbf{Q} & \mathbf{1}_k \end{pmatrix}$ has full column rank (e.g., Lang, 1987, pp. 59-60). Here we propose a different approach.

In Section 2, it is explained how the $\mathbf{Q}$-matrix is adapted to impose a convenient normalization. Neither the normalization nor the particular parameterization of the basic parameters has an effect on the differences between the item parameters or on the fit of the model, so that there are infinitely many ways to specify the same LLTM. In Section 3 we will discuss such equivalent LLTMs. An interesting implication is that researchers may think differently about a given set of items and nevertheless specify statistically equivalent models. Another implication is that there may well be many ways to change the specification of a given LLTM and achieve the same improvement in model fit. The latter implication will be illustrated in Section 4 where we discuss a test for the hypothesis that a single entry in the **Q**-matrix is specified correctly. In line with Section 4, Section 5 is about the estimation of this entry when it is specified as a parameter. In Section 6, we analyze a real data set. Finally, we summarize and discuss our findings in Section 7.

2 Normalizing the LLTM

As noted in the Introduction, researchers usually specify the **Q**-matrix for k item parameters, and ignore the need for a normalization. As a first step in an analysis with the LLTM, one should therefore make sure that a normalization is imposed. Let $\mathbf{I}_k$ denote a k-dimensional unit matrix. Consider an arbitrary normalization indexed by $\mathbf{a} = (a_1, ..., a_k)^t$. Given a vector of item parameters, the **a**-normalized item parameters are

$$\boldsymbol{\beta}_{\mathbf{a}} = \mathbf{L}_{\mathbf{a}}\boldsymbol{\beta}, \tag{6}$$

where $\mathbf{L}_{\mathbf{a}} = \mathbf{I}_k - \mathbf{1}_k\mathbf{a}^t$ and $\mathbf{a}^t\mathbf{1}_k = 1$. Substituting the postulated LLTM for the item parameters gives

$$\boldsymbol{\beta}_{\mathbf{a}} = \mathbf{L}_{\mathbf{a}}\mathbf{Q}\boldsymbol{\eta} = \mathbf{Q}_{\mathbf{a}}\boldsymbol{\eta}. \tag{7}$$

Suppose, for example, that an LLTM is formulated as

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}. \tag{8}$$

The differences between any two item difficulty parameters can be arranged in the following matrix

$$\begin{bmatrix} 0 & -\eta_1 + 2\eta_2 & \eta_2 \\ \eta_1 - 2\eta_2 & 0 & \eta_1 - \eta_2 \\ -\eta_2 & -\eta_1 + \eta_2 & 0 \end{bmatrix}, \tag{9}$$

where the (i, j) entry is the difference between β_i and β_j. This matrix may be used to specify the LLTM for differences between the item parameters. Its first column, for instance, provides the model for differences with the first item. This is an LLTM with the following "normalized" **Q**-matrix:

$$\mathbf{Q}_{(1,0,0)^t} = \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ 0 & -1 \end{bmatrix}. \tag{10}$$

By adapting the **Q**-matrix we have obtained the model for difficulties of the items relative to the first item.

The second step in the analysis is to check the rank of the "normalized **Q**-matrix", $\mathbf{Q_a}$, to make sure that the basic parameters are identified.

Proposition 2 *The basic parameters are identified if and only if* $\mathbf{Q_a}$ *has full column rank.* ***Proof.*** *The item parameters are identified because we have applied a normalization. That is,* $\boldsymbol{\beta}_{\mathbf{a}}^{(1)} \neq \boldsymbol{\beta}_{\mathbf{a}}^{(2)} \Rightarrow \Pr(X_{vi} = x_{vi}|\xi_v, \beta_i^{(1)}) \neq \Pr(X_{vi} = x_{vi}|\xi_v, \beta_i^{(2)})$*, for some* $i = 1, ..., k$*. Hence, we must establish that* $\boldsymbol{\eta}^{(1)} \neq \boldsymbol{\eta}^{(2)} \Rightarrow \boldsymbol{\beta}_{\mathbf{a}}^{(1)} \neq \boldsymbol{\beta}_{\mathbf{a}}^{(2)}$*. This is well known to apply if and only if* $\mathbf{Q_a}$ *has full column rank (e.g., Lang, 1987, pp. 59-60).* ■

The rank of $\mathbf{Q_a}$ is independent of the normalization. It is easily checked that $\mathbf{L_a L_b} = \mathbf{L_a}$, where **b** and **a** denote two normalizations. Since $\text{rank}(\mathbf{L_a Q}) = \text{rank}(\mathbf{L_a L_b Q}) \leq \text{rank}(\mathbf{L_b Q})$ while $\text{rank}(\mathbf{L_b Q}) = \text{rank}(\mathbf{L_b L_a Q}) \leq \text{rank}(\mathbf{L_a Q})$ it follows that $\text{rank}(\mathbf{L_a Q}) = \text{rank}(\mathbf{L_b Q})$.

3 Equivalent LLTMs

There are many equivalent formulations of a LLTM because we can change the normalization and/or the meaning of the basic parameters without changing the restrictions on the differences between the item parameters. As explained in the previous section, the normalization may be changed by means of the $\mathbf{L_a}$ matrix. The meaning of the basic parameters is changed by means of a non-singular transformation defined by $\boldsymbol{\eta}^* = \mathbf{T}\boldsymbol{\eta}$, where $\mathbf{T}$ is a non-singular matrix. In general, therefore, an LLTM may be written as :

$$\boldsymbol{\beta}_{\mathbf{a}} = \mathbf{L_a Q T}^{-1}\mathbf{T}\boldsymbol{\eta} \equiv \mathbf{Q_a T}^{-1}\mathbf{T}\boldsymbol{\eta}, \tag{11}$$

where different choices of **a** and **T** give different specifications of the same LLTM.

An interesting implication is that researchers who hold different theories about a given set of items may nevertheless specify LLTMs that are equivalent from a mathematical point of view. How do we determine whether two LLTMs are equivalent ?

Definition 3 *In general, two LLTMs* $\mathbf{Q}^{(1)}\boldsymbol{\eta}^{(1)}$ *and* $\mathbf{Q}^{(2)}\boldsymbol{\eta}^{(2)}$ *are equivalent if*

$$\mathbf{Q}_{\mathbf{a}}^{(1)}\boldsymbol{\eta}^{(1)} = \mathbf{Q}_{\mathbf{a}}^{(2)}\mathbf{T}^{-1}\mathbf{T}\boldsymbol{\eta}^{(2)}. \tag{12}$$

If $\mathbf{Q}_{\mathbf{a}}^{(1)}$ is in the column-space of $\mathbf{Q}_{\mathbf{a}}^{(2)}$, we may always find a non-singular matrix $\mathbf{T}^{-1}$ such that $\mathbf{Q}_{\mathbf{a}}^{(1)} = \mathbf{Q}_{\mathbf{a}}^{(2)}\mathbf{T}^{-1}$. Hence, two LLTMs are equivalent when both $\mathbf{Q}_{\mathbf{a}}$-matrices span the same column space, and the rank of the augmented matrix $[\mathbf{Q}_{\mathbf{a}}^{(1)}; \mathbf{Q}_{\mathbf{a}}^{(2)}]$ is p. The transformation matrix may then be calculated as:

$$\mathbf{T}^{-1} = \left(\mathbf{Q}_{\mathbf{a}}^{(1)}\right)^{g} \mathbf{Q}_{\mathbf{a}}^{(2)}, \tag{13}$$

where $\left(\mathbf{Q}_{\mathbf{a}}^{(1)}\right)^{g}$ is an arbitrary g-inverse of $\mathbf{Q}_{\mathbf{a}}^{(1)}$ (see e.g., Pringle & Rayner, 1971).

Another implication is that the same improvement in the specification of a postulated **Q**-matrix can be achieved in many ways. As an illustration we will, in the next section, derive an efficient score (Rao, 1947) or Lagrange multiplier test (Aitchison & Silvey, 1958; Silvey, 1959) for the hypothesis that an arbitrary entry in the **Q**-matrix is specified correctly.

4 A Lagrange Multiplier Test for the LLTM

Suppose we consider the postulated LLTM as the null-hypothesis. To test the specification of the (g, l) entry in the **Q**-matrix, q_{gl}, we compare the postulated LLTM to an alternative model where q_{gl} is specifies as a parameter σ_{gl}. That is,

$$H_0 \; : \; \sigma_{gl} = q_{gl} \tag{14}$$

$$H_1 \; : \; \sigma_{gl} \neq q_{gl} \tag{15}$$

If the null hypothesis is true, an estimate of σ_{gl} would be near q_{gl}.[1] We could use a likelihood ratio test but this requires estimation under the alternative model. To avoid this, we employ the Lagrange Multiplier (LM) test. The LM test is similar to the likelihood ratio test but it does not require estimation under the general model (see Buse, 1982).

Let $\boldsymbol{\theta} = (\boldsymbol{\eta}^t, \sigma_{gl})^t$ denote the vector of parameters for the general model. Let $\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0)$ be the vector of first order partial derivatives of the conditional loglikelihood (i.e., the score vector) of the general model evaluated at $\boldsymbol{\theta}_0$, which denotes the Conditional Maximum Likelihood (CML) estimates under the LLTM. That is, $\boldsymbol{\theta}_0 = (\hat{\boldsymbol{\eta}}^t, q_{gl})^t$, where $\hat{\boldsymbol{\eta}}$ denotes the CML estimate of the basic parameters. Similarly, $\mathbf{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\theta}_0)$ denotes the expected conditional information matrix evaluated at $\boldsymbol{\theta}_0$. Since $\boldsymbol{\theta} = (\boldsymbol{\eta}^t, \sigma_{gl})^t$ is partitioned, the score vector and the information matrix may likewise be partitioned:

$$\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{S}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_0) \\ \mathbf{S}_{\sigma_{gl}}(\boldsymbol{\theta}_0) \end{pmatrix}, \mathbf{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{I}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\boldsymbol{\theta}_0) & \mathbf{I}_{\boldsymbol{\eta},\sigma_{gl}}(\boldsymbol{\theta}_0) \\ \mathbf{I}_{\sigma_{gl},\boldsymbol{\eta}}(\boldsymbol{\theta}_0) & \mathbf{I}_{\sigma_{gl},\sigma_{gl}}(\boldsymbol{\theta}_0) \end{pmatrix}. \tag{16}$$

The LM test involves the following statistic (cf. Glas & Verhelst, 1995):

$$\varphi_{gl} = S_{\sigma_{gl}}(\boldsymbol{\theta}_0)^2 D_{gl}^{-1}, \tag{17}$$

where

$$D_{gl} = \mathbf{I}_{\sigma_{gl},\sigma_{gl}}(\boldsymbol{\theta}_0) - \mathbf{I}_{\sigma_{gl},\boldsymbol{\eta}}(\boldsymbol{\theta}_0)\mathbf{I}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\boldsymbol{\theta}_0)^{-1}\mathbf{I}_{\boldsymbol{\eta},\sigma_{gl}}(\boldsymbol{\theta}_0).$$

If the postulated LLTM holds, this statistic follows a chi-square distribution with one degree of freedom. Another view on the LM is given by Sörbom (1989), who demonstrates that $\frac{1}{2}\varphi_{gl}$ is approximately the increase in conditional loglikelihood when the corresponding element of the **Q**-matrix is estimated (or specified as a constant with the value of the estimate).

Note that

$$S_{\sigma_{gl}}(\boldsymbol{\theta}_0)\hat{\eta}_l^{-1} = E[x_{.g}|\mathbf{R} = \mathbf{r}] - x_{.g} \quad (g \in \{1, ..., k\}, l \in \{1, ..., p\}) \tag{18}$$

provides an informal measure of the fit of item g under the LLTM. The character $x_{.g}$ denotes the number of correct responses to item g, and $E[x_{.g}|\mathbf{R} = \mathbf{r}]$ denotes its expectation conditional on the observed subject scores. For later reference, we call (18) the *Item Fit (IF) index.*

[1] Note that the alternative model pertains to a class of models that is more general than the LLTM. A definition of these models and conditions for their identifiability are given by Bechger, Verhelst and Verstralen (2000).

It can be shown that φ_{gl} is invariant under normalization and transformations of the basic parameters. This means that φ_{gl} and φ_{jh} (where $g, j \in \{1, ..., k\}$, and $l, h \in \{1, ..., p\}$) will be the same if the corresponding general models are equivalent. The reverse implication does not hold because φ_{gl} and φ_{jh} may be equal by change.

Theorem 4 *Let φ_{gl} and φ_{jh} denote the statistics associated to two entries in a given* **Q***-matrix. Let $M^{(1)}$ denote a model where only the (g, l) entry in* **Q** *is estimated as a parameter, and let $M^{(2)}$ denote a model where only the (j, h) entry is estimated as a parameter. Then, $\varphi_{gl} = \varphi_{jh}$ if $M^{(1)}$ is equivalent to $M^{(2)}$. The proof of this theorem is added as an appendix.*

This finding is intuitively plausible. When the same improvement in conditional loglikelihood can be achieved by estimating (or re-specifying) one of several entries in the **Q**-matrix, the corresponding statistics should indeed be equal.

5 Estimating σ_{gl}

Suppose we keep the basic parameters at the CML estimates and consider Fisher's Scoring (FS) algorithm to find optimal values for the parameter in the **Q**-matrix. After simplification of the gradient and the information matrix we find the following relation between estimates at iteration t and $t + 1$:

$$\sigma_{gl}^{(t+1)} = \sigma_{gl}^{(t)} + D^{-1} S_{\sigma_{gl}}(\boldsymbol{\theta}_0), \tag{19}$$

where $D^{-1} S_{\sigma_{gl}}(\boldsymbol{\theta}_0)$ is evaluated at the current values of the basic parameters. If we start the iterations at $\sigma_{gl}^{(0)} = q_{gl}$, the quantity $D^{-1} S_{\sigma_{gl}}(\boldsymbol{\theta}_0)$, which equals φ_{gl} divided by $S_{\sigma_{gl}}(\boldsymbol{\theta}_0)$, gives $\sigma_{gl}^{(1)} - q_{gl}$. In the next section, we call this quantity the *First FS Step (FFSS)*. To estimate the optimal value of q_{gl}, we fix q_{gl} at the estimated value of σ_{gl}, estimate the basic parameters, take the next FS step, and continue this procedure until the specification of q_{gl} can no longer be improved. We have then replaced q_{gl} by a CML estimate of the parameter σ_{gl}, given the values of the other entries in **Q**. An estimate of its standard error is given by the square root of the inverse of $\mathbf{I}_{\sigma_{gl}, \sigma_{gl}}(\boldsymbol{\theta})$ evaluated on the final estimates. Note that the estimates of an entry in **Q** will be biased when other entries of **Q** are also misspecified.

This procedure makes sense if σ_{gl} is identified and its value is unique. The following theorem provides a relatively simple way to determine the identifiability of single entries in the **Q**-matrix. It is proven in Bechger, Verhelst and Verstralen (2000).

Theorem 5 *Choose a normalization such that* $a_g = 0$ *and let* $\mathbf{Q}_{\mathbf{a}(-g)}$ *denote* $\mathbf{Q}_{\mathbf{a}}$ *with the* g*th row deleted. Then, the* (g, l) *entry in* $\mathbf{Q}$ *is identifiable if and only if* $\mathbf{Q}_{\mathbf{a}(-g)}$ *has full column rank and* $\eta_l \neq 0$.

As an illustration, consider the following LLTM:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 1 & \sigma_{32} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \eta_1 \\ 2\eta_1 + \eta_2 \\ \eta_1 + \sigma_{32}\eta_2 \end{pmatrix}, \tag{20}$$

where the $(3, 2)$ entry is considered a parameter. Suppose that we do not normalize the LLTM. If we solve the equations for σ_{32} we find that

$$\sigma_{32} = \frac{\beta_3 - \beta_1}{\beta_2 - 2\beta_1}. \tag{21}$$

One might be tempted to conclude that σ_{32} is identified but this is not true because $\beta_2 - 2\beta_1$ depends on the normalization; that is, if $c \neq 0$, $\beta_2 - 2\beta_1 \neq (\beta_2 + c) - 2(\beta_1 + c)$. This shows how important it is to consider the normalized **Q**-matrix rather than the matrix a researcher provides us with. The conclusion that σ_{32} is not identified is confirmed if we look at the rank of the $\mathbf{Q}_{\mathbf{a}(-3)}$ matrix. In the example that was mentioned in Section 2,

$$\mathbf{Q}_{\mathbf{a}(-2)} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, \tag{22}$$

and σ_{21} is not identified. Both examples illustrate that the rank of $\mathbf{Q}_{\mathbf{a}(-g)}$ is at most $k-2$ so that no element of the **Q**-matrix will be identified if $p > k-1$. When $p < k-1$, the **Q**-matrix may contain entries that are identified as well as entries that are not identified.

Note that if we consider η_l to be random with some non-degenerate, continuous distribution, the probability that $\eta_l = 0$ is zero. If, however, in applications $|\eta_l|$ becomes very small, the FS algorithm may fail to converge.

6 An Application

Consider as an illustration a data set consisting of 300 responses to five items that is published by Rost (1996, e.g., p. 303). For these five items, the following **Q**-matrix was believed to be appropriate (Rost, 1996, section 3.4.1):

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 1 & 1 \\ 2 & 1 \\ 2 & 2 \end{pmatrix}. \tag{23}$$

The entries in this **Q**-matrix specify the number of times that each of two cognitive operations is required to solve the corresponding item.

We have used CML estimation as described briefly in the Appendix. The basic parameters (with their asymptotic standard error within parenthesis) are estimated to be $\hat{\eta}_1 = 0.46$ (0.14), and $\hat{\eta}_2 = 0.96$ (0.10). The conditional loglikelihood of the model is -324.92, which is only slightly lower than the conditional loglikelihood for the RM, which is -320.62. (Using a central chi-square approximation, minus twice this difference is associated with a probability level of 0.01.) As judged from the IF indices (see Equation 18), item 4 and item 5 are the worst fitting items. Item 4, for example, is associated to an IF index of 14.35, which indicates that there are about 14 correct responses to many under the postulated LLTM. Since this is only a small percentage of the sample, the IF index suggest that the model fits the data very well.

The φ's are

$$\begin{pmatrix} 1.61 & 1.61 \\ 1.76 & 1.76 \\ 1.61 & 1.61 \\ 6.93 & 6.93 \\ 8.30 & 8.30 \end{pmatrix}. \tag{24}$$

Equalities between the MIs suggest "equivalent" ways to improve the model. For starters, the MIs are equal for each row of **Q** since an equivalent improvement in loglikelihood can be achieved by a change in either one of the weights for the first or the second basic parameter. To be more precise, if the fit of the LLTM can be improved by adding a constant d to the estimated

β_i, this improvement can be achieved in two ways since

$$\hat{\beta}_i + d = (q_{i1} + \frac{d}{\eta_1})\eta_1 + q_{i2}\eta_2 = q_{i1}\eta_1 + (q_{i2} + \frac{d}{\eta_2})\eta_2. \quad (25)$$

It is obvious that this phenomenon will appear in any LLTM, even when there are more than two basic parameters. The equality between the MIs of items 1 and 3 can also be explained by the presence of equivalent models. To be more specific, we found that an LLTM with the $(1,2)$ entry specified as -0.347 is equivalent to an LLTM with entry $(1,2)$ fixed at 0 and entry $(3,2)$ specified as $1 + 0.347$. The value 0.347 was found using the FS procedure described in the previous section. With either specification, the conditional loglikelihood improved $\frac{1}{2} \times 1.61$.

The results suggests that the greatest improvement in conditional loglikelihood (about 4.15) can be achieved by estimating either entry in the fifth row. Since the entries in the **Q**-matrix represent the number of times that a certain cognitive operation is used in the solution process, it seems reasonable to consider only integer values. The FFSS suggests that the $(5,1)$ entry should be about 2 lower and that the $(5,2)$ entry should be about 1 lower. If we specify $q_{51} = 0$ we obtain an LLTM with a loglikelihood of -321.15. The basic parameters are now estimated to be $\hat{\eta}_1 = 0.49$ (0.09), and $\hat{\eta}_2 = 1.38$ (0.10). Suppose that instead of changing the value of q_{51}, we change the value of q_{41} to 1. The φ's for the parameters of the fifth item are no longer significant and the fit of the model has improved to -321.68, which is only slightly lower than the likelihood of the model with $q_{51} = 0$. This shows that the φ's are not independent. The basic parameters are now estimated to be $\eta_1 = 0.38$ (0.09) and $\eta_2 = 0.95$ (0.10), close to the values under the first model. Although the different specifications are not equivalent, substantial arguments are needed to distinguish between them.

7 Conclusion

Researchers often specify the **Q**-matrix as if there were k independent item difficulty parameters and ignore the need for a normalization. To deal with this unfortunate situation, Fischer (e.g., 1995) suggests that we examine the **Q**-matrix to assure ourselves that a normalization is imposed. Unlike Fischer, we have proposed that a given **Q**-matrix be adapted to make sure that a convenient normalization is imposed. We have found that the basic parameters are identified if and only if the adapted **Q**-matrix is of full column rank. If the adapted **Q**-matrix is of deficient column rank, the researcher is required to specify a new **Q**-matrix.

We have also found that there are infinitely many ways to specify an LLTM because one is completely free to choose a normalization or transform the basic parameters. An interesting implication is that researchers may think differently about a given set of items and nevertheless specify statistically equivalent models. Another implication is that there are many ways to change the specification of a given LLTM and achieve the same improvement in model fit. We have demonstrated that this affects the LM test for single entries in the **Q**-matrix. In general, the statistics associated with different sets of entries should be equal if the same improvement can be achieved by changing the specification of either set of entries.

It may be tempting to use the LM test in the LLTM in the same way as is done in Structural Equation Modeling (SEM). That is, to examine the statistics, to free the restriction that leads to the largest reduction in goodness-of-fit and to repeat the process with the revised model until an adequate fit is developed. Bollen (1989) surveys years of experience with the LM test and puts in the following caveats against this procedure:

1. The central chi-square distribution may not even be valid for statistics associated with correctly specified entries. The reason is that the wrong values in other places in the **Q**-matrix are part of the null hypothesis.

2. The statistics are not independent.

3. The order in which parameters are freed or restricted can affect the significance tests for the remaining parameters.

4. The LM test is most useful when the misspecifications are minor. Fundamental changes in the structure of the model may be needed but these are not detectable with these procedures.

5. Simulation studies show that a specification search that is guided on the LM test is quite likely to lead one to the wrong model.

These findings highlight the importance of guidance from substantive knowledge in the specification of the LLTM and suggest that the LM test should be used with caution. Finally, we note that it is common practice in SEM to supplement the LM test with a preliminary estimate of the parameter, which is given by the first FS step (see Section 5).

8 Appendix:

8.1 The LM test depends neither on the normalization, nor on the parameterization of the basic parameters

Theorem 6 *The value of the statistic associated to the LM test depends neither on the normalization, nor on the parameterization of the basic parameters.*

To prove this statement we proceed in two steps. First, we prove that the statistic φ_{gl} is independent of the normalization. Then, we prove that the value of the statistic is not changed by a transformation of the basic parameters. We start by giving general expressions for the score vector and the conditional information matrix.

In general, the statistic φ is given by

$$\varphi = \mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0)^t \mathbf{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\theta}_0)^{-1} \mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0). \tag{26}$$

It follows from the chain rule that $\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is given by

$$\mathbf{J}^t \left(\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right), \tag{27}$$

where $l(\boldsymbol{\beta})$ denotes the conditional loglikelihood of the item parameters under the RM, $\mathbf{J} = (\frac{\partial \beta_i}{\partial \theta_j})$ denotes the Jacobian matrix of first-order partial derivatives (e.g., 35), and θ_j denotes a parameter is $\boldsymbol{\theta} = (\boldsymbol{\eta}^t, \sigma_{gl})^t$. (Bechger, Verhelst and Verstralen (2000) give a general expression for this Jacobian matrix.) The conditional information matrix is given by

$$\mathbf{I}_{\boldsymbol{\theta},\boldsymbol{\theta}}(\boldsymbol{\theta}) = \mathbf{J}^t \left(-E \left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \right] \right) \mathbf{J}. \tag{28}$$

The derivatives with respect to the item parameters can be found, e.g., in Fischer (1974 or 1995).

Suppose that we change to a different normalization indexed by $\mathbf{a}$. The Jacobian matrix is now:

$$\mathbf{J}_{\mathbf{a}} = \mathbf{L}_{\mathbf{a}} \mathbf{J} = \mathbf{J} - \mathbf{1}_k \mathbf{a}^t \mathbf{J}. \tag{29}$$

We use the following identities, and substitute $\mathbf{J_a}$ in the equation for the statistic φ to obtain $\varphi_{\mathbf{a}}$; the LM test statistic under arbitrary normalization.

$$\left(-E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^t}\right]\right)\mathbf{1}_k = \mathbf{0}_k, \text{ and } \left(\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)^t \mathbf{1}_k = 0 \tag{30}$$

The first identity is proven by Verhelst (1993, lemma 1). The second identity follows easily by substituting the derivatives. Now the LM test statistic is

$$\varphi_{\mathbf{a}} = \left(\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)^t \mathbf{L_a J}\left[\mathbf{J}^t\mathbf{L}_{\mathbf{a}}^t\left(-E\left[\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^t}\right]\right)\mathbf{L_a J}\right]^{-1}\mathbf{J}^t\mathbf{L}_{\mathbf{a}}^t\left(\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right) \tag{31}$$

If one substitutes (29) in this equation, and uses the identities (30), one finds that $\varphi_{\mathbf{a}} = \varphi$. This proofs that the statistic φ is independent of the normalization. It follows from the chain rule, and the rule about the transformation of the information after reparameterization (e.g., Azzalini, 1996, Equation 3.11) that the statistic φ is also invariant under various parameterizations of the basic parameters. Thus, if two alternative hypotheses are equivalent, the associated statistics will be the same. ■

8.2 A Common Framework for CML Estimation of the RM and the LLTM

In this paragraph we demonstrate that CML estimates of the item parameters in the RM or the basic parameters in the LLTM can be obtained using the same algorithm. This algorithm may not efficient because it involves multiplication with matrices that may in practice be very large and sparse but may be useful with small data sets (as in the present paper) or in theoretical work.

First, the researcher chooses a $\mathbf{Q}$-matrix and a normalization, i.e., a vector $\mathbf{a}$, such that $\mathbf{a}^t\mathbf{1}_k = 1$. The CML estimates of the normalized item parameters are

$$\hat{\boldsymbol{\beta}}_{\mathbf{a}} = \mathbf{L_a Q}\boldsymbol{\eta} = \mathbf{Q_a}\hat{\boldsymbol{\eta}}. \tag{32}$$

where $\hat{\boldsymbol{\eta}}$ denotes the CML estimates of the basic parameters and $\mathbf{L_a} = \mathbf{I}_k - \mathbf{1}_k\mathbf{a}^t$. The asymptotic variance-covariance matrix of the normalized item parameters is

$$\mathbf{L_a Q I}_{\eta,\eta}^{-1}(\hat{\eta})\mathbf{Q}^t\mathbf{L}_{\mathbf{a}}^t. \tag{33}$$

where $\mathbf{I}^{-1}_{\boldsymbol{\eta},\boldsymbol{\eta}}(\hat{\boldsymbol{\eta}})$ denotes the asymptotic variance-covariance matrix of the estimates of the basic parameters.

If rank$(\mathbf{Q_a}) = p$ the basic parameters are identified and we may use a FS algorithm to estimate the basic parameters. The FS algorithm requires the score vector and information matrix, which are given by (27) and (28), respectively. The Jacobian matrix $\mathbf{J} = \mathbf{L_a Q}$. The FS algorithm involves the evaluation of the elementary symmetric functions which is described by Verhelst, Glas and Van der Sluis (1984), or Verhelst and Veldhuijzen (1991).

To fit the RM, we may choose $\mathbf{Q} = \mathbf{N}_r$, which is a $k \times k-1$ matrix that we obtain by deleting the rth row of $\mathbf{I}_k$. Note that if $\mathbf{Q} = \mathbf{N}$, $\boldsymbol{\eta}^t = (\beta_a - \beta_r, \beta_{a-1} - \beta_r, ..., \beta_{r-1} - \beta_r, \beta_{r+1} - \beta_r, ..., \beta_{r+b} - \beta_r)$, where

$$a = \begin{cases} 2 \text{ if } r = 1 \\ 1 \text{ otherwise} \end{cases} \tag{34}$$

and $b = k - r$. That is, the CML estimates must be interpreted as distances relative to the rth item.

To calculate the LM statistic for the (g, l) entry in the $\mathbf{Q}$-matrix it is convenient to choose a normalization where $a_g = 0$. The Jacobian matrix is then given by

$$\mathbf{J} = [\mathbf{L_a Q}(\sigma_{gl}), \mathbf{P}(\eta_l)], \tag{35}$$

where $\mathbf{P}(\eta_l)$ denotes a column vector with all elements equal to zero except the gth entry, which is equal to η_l. The notation $\mathbf{Q}(\sigma_{gl})$ indicates that $\mathbf{Q}$ is now a function of the parameter σ_{gl}. The score vector and the conditional information matrix are again given by (27) and (28), respectively.

9 References

Aitchison, W., & Silvey, S. D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, *28*, 813-828.

Azzalini, A. (1996). *Statistical inference based on the likelihood.* London: Chapman and Hall.

Bechger, T. M. , Verhelst, N. D. , & Verstralen, H. (to appear). Identifiability of non-linear logistic test models. Accepted for Publication in Psychometrika.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier test: An expository note. *The American Statistician*, *36* (3), 153-157.

Fischer, G. H. (1974). *Einfurung in die Theorie psychologischer Tests: Grundlagen und Anwendungen.* [Introduction to the theory of psychological tests: Foundations and applications.] Bern: Verlag Hans Huber.

Fischer, G. H. (1995). The linear logistic test model. Chapter 8 in G. H. Fischer and I.W. Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications.* Berlin: Springer.

Glas, C. A. W. , & Verhelst, N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Their foundations, recent developments and applications.* New-York: Springer Verlag.

Lang, S. (1987). *Linear Algebra.* (Third Edition). New-York: Springer.

Pringle, R. M. , & Rayner, A. A. (1971). *Generalized inverse matrices with applications to Statistics.* London: Charles Griffin.

Rao, C. R. (1947). Large sample tests of statistical hypothesis concerning several parameters with applications to the problems of estimation. *Proceedings Cambridge Philosophical Society*, *44*, 50-57.

Rost, J. (1996). *Testtheory, Testkonstruction* [Test theory, test construction]. Bern: Verlag Hans Huber.

Silvey, S. D. (1959). The Lagrange multiplier test. *Annals of Mathematical Statistics*, *30*, 398-407.

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371-384.

Verhelst, N. D. (1993). *On the standard errors of parameter estimators in the Rasch model.* (Measurement and Research department Reports, 93-1.). Arnhem, Cito.

Verhelst, N. D., Glas, C. A. W., & Van der Sluis, A. (1984). Estimation problems in the Rasch model: The basis symmetric functions. Computational Statistics Quarterly, 245-262.

Verhelst, N. D., Veldhuijzen, N. H. (1991). A new algorithm for computing elementary symmetric functions and their first and second order derivatives. (Measurement and Research Department Reports 91-5). Arnhem: Cito.