Psychology Science, Volume 49, 2007 (1), p. 26-43

CML based estimation of extended Rasch models with the eRm package in R

PATRICK MAIR[1] & REINHOLD HATZINGER

Abstract

This paper presents an open source tool for computing extended Rasch models. It is realized in `R` (R Development Core Team, 2006) and available as package `eRm`. In addition to ordinary Rasch models extended models such as linear logistic test models, (linear) rating scale models and (linear) partial credit models can be estimated. A striking feature of this package is the implementation of conditional maximum likelihood estimation techniques which relate directly to Rasch's original concept of specific objectivity. The mathematical and epistemological benefits of this estimation method are discussed. Moreover, the capabilities of the `eRm` routine with respect to structural item response designs are demonstrated.

Key words: Extended Rasch models, CML estimation, specific objectivity, eRm-package.

[1] Correspondence regarding this article may be addressed to Patrick Mair, Department of Statistics and Mathematics, Vienna University of Economics, Augasse 2-6, 1090 Vienna. Email: patrick.mair@wu-wien.ac.at; reinhold.hatzinger@wu-wien.ac.at

1. Introduction

Item response theory (IRT) models have a long tradition in psychological testing. The most prominent model is the Rasch model formulated by Rasch in 1960. In the aftermath, various generalizations have been developed such as the *linear logistic test model* (LLTM; Scheiblechner, 1972; Fischer, 1973), the *rating scale model* (RSM; Andrich, 1978), the *linear rating scale model* (LRSM; Fischer & Parzer, 1991), the *partial credit model* (PCM; Masters, 1982), and the *linear rating scale model* (LPCM; Fischer & Ponocny, 1994). These models will be presented briefly in Section 2. A more detailed overview can be found in Fischer and Molenaar (1995), Mair and Hatzinger (2007), Rost (1988), and Rost and Straß (1992). Additional IRT models are presented in Embreston and Reise (2000), and Baker and Kim (2004).

A common feature of the listed model family is that all of them can be estimated using conditional maximum likelihood (CML) methods. As will be described in Section 3, in addition to its persuasive mathematical properties this estimation technique is well founded from an epistemological point of view.

A general problem in IRT modeling is the computational implementation in terms of software packages. Most of the packages are commercially distributed and thus the general availability is restricted. Additionally, the usability and the handling of some programs are not very transparent. Other programs are too restricted onto simple models and do not provide flexible issues like polytomous or longitudinal IRT models with possible treatment effects. The motivation for developing the `eRm` package (Mair & Hatzinger, 2006) was to provide an open source software which allows for the computation of various CML-based IRT models and to provide a flexible and usable interface to compute polytomous Rasch models with time and treatment effect. More features will be described in Section 4.

2. Extended Rasch models

The family of Rasch models implemented in the `eRm` package have the following common properties: Unidimensionality of the latent trait, sufficiency of the raw scores, parallel item characteristic curves (ICC), and local independence. For these models we introduced the terminus „extended Rasch models" as an umbrella term.

Let $\mathbf{X}$ be a data matrix made up by the responses of $v = 1,\ldots,n$ subjects to $i = 1,\ldots,k$ items. Accordingly, the response of subject v on item i is denoted as x_{vi}. The basic model developed by Rasch (1960, 1980) is

$$P(X_{vi} = 1) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}, \tag{1}$$

where β_i is the item parameter and θ_v the subject parameter. In the context of psychological testing, β_i is usually considered as item difficulty and θ_v as a person's ability. Basically, each item parameter can be reformulated by the linear combination

$$\beta_i = \sum_{j=1}^{p} w_{ij}\eta_j \ , \tag{2}$$

where η_j represents the so called "basic parameter". The corresponding weights w_{ij} have to be fixed a priori and form the design matrix **W** of dimension $(k \times p)$. Thus, the general form of Equation 2 is $\boldsymbol{\beta} = \mathbf{W}\boldsymbol{\eta}$. Equation 2 can be seen from two different angles: On the one hand this expression can be considered as a more parsimonious reparameterization of Equation 1 whereas on the other hand it can be regarded as a generalization of the basic Rasch model. For example, Scheiblechner (1972) used the (resulting) LLTM as a more parsimonious Rasch model to analyze a test on a set operation skills, i.e., $k > p$, such that each item parameter β_i was composed by some other parameters describing "cognitive operations" η_j (e.g., negation, disjunction, conjunction etc.). If $k < p$, the LLTM can be used as a linearly extended Rasch model by introducing the concept of *virtual items* in terms of time and treatment effects as given in Fischer (1995b).

The Rasch model in its original form is limited to dichotomous items. For polytomous items with the same number $h = 0,\ldots,m$ of response categories per item, Andrich (1978) proposed the RSM

$$P(X_{vi} = h) = \frac{\exp[h(\theta_v + \beta_i) + \omega_h]}{\sum_{l=0}^{m} \exp[l(\theta_v + \beta_i) + \omega_l]} \ , \tag{3}$$

where the category parameters $\omega_0,\ldots,\omega_m$ describe the scoring which is considered to be the same for all items.

For cases where the constant scoring property does not hold and/or the number of categories differs across the items, i.e., m_i instead of m, the PCM as proposed by Masters (1982) may be appropriate, i.e.,

$$P(X_{vih} = 1) = \frac{\exp[h\theta_v + \beta_{ih}]}{\sum_{l=0}^{m_i} \exp[l\theta_v + \beta_{il}]} \ , \tag{4}$$

where the β_{ih}'s describe item-category combinations.

In analogy to the relation between the Rasch model and the LLTM, the linear extension of the β-parameters due to Equation 2 can be established for the RSM and the PCM as well. The resulting models are the LRSM (Fischer & Parzer, 1991) and the LPCM (Fischer & Ponocny, 1994).

It is obvious that these models are hierarchically nested where the LPCM is the most general model and all other models are special cases (see Figure 1). This issue is important since once the likelihood expressions for the LPCM are established all other models can be estimated using the same equations. The corresponding likelihood equations and first order derivatives as well as some discussion on computational aspects are given in Fischer and Ponocny (1994), the second order derivatives can be found in Mair and Hatzinger (2007). A general description of the CML estimation method is given in the following section.

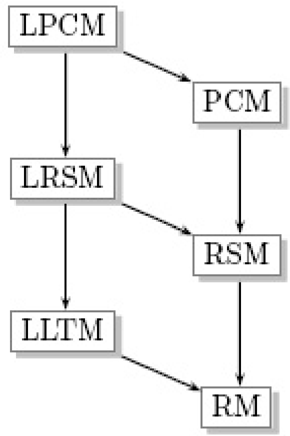

Figure 1: Model hierarchy

3. CML estimation and the relation to specific objectivity

As mentioned in the introduction, the `eRm` package is limited to the family of extended Rasch models where CML estimation is applicable. Very general expressions of such CML-compatible models are provided by Andersen (1995):

$$P(X_{vi}=h)=\frac{\exp[\phi_h(\theta_v+\beta_i)+\omega_h]}{\sum_{l=0}^{m_i}\exp[\phi_l(\theta_v+\beta_i)+\omega_l]} \tag{5}$$

and

$$P(X_{vi}=h)=\frac{\exp[\phi_h\theta_v+\beta_{ih}]}{\sum_{l=0}^{m_i}\exp[\phi_l\theta_v+\beta_{il}]}. \tag{6}$$

In the above equations, ϕ_h is the scoring function for category h. For the RSM, the LRSM, the PCM, and the LPCM $\phi_h = h$ whereas for the remaining models $\phi_h = 1$.

It should be mentioned that already Rasch (1961) presented a polytomous generalization where, e.g, the well known 2-PL model (Birnbaum, 1968) fits into. However, for this general expression CML estimation is not possible, since the crucial condition for the use of CML is not fulfilled, i.e., sufficiency of the raw score. In other words, CML requires a person's raw score $r_v = \sum_{i=1}^{k} x_{vi}$ to be a minimal sufficient statistic for θ_v.

In the following sections we restrict our considerations to the basic Rasch model (except for one LLTM example). The reason is that the dichotomous Rasch model and its (dichotomous) extensions such as the LLTM and the LLRA (*linear logistic test model with relaxed assumptions*; Fischer, 1977) are the only models which can be derived directly from the assumption of *specific objectivity* (see Fischer 1987, 1995a). Note that Irtel (1995) provides an extension to this concept allowing for additional models such as the 2-PL to be regarded within the framework of specific objectivity. However, a thorough treatment is beyond the scope of this paper.

3.1 Mathematical properties of the CML estimates

A variety of estimation approaches for IRT models in general and for the Rasch model in particular are available: the *joint maximum likelihood* (JML) estimation as proposed by Wright and Panchapakesan (1969) which is not recommended since the estimates are not consistent (see, e.g., Haberman, 1977). The basic reason for that is that the person parameters θ are nuisance parameters; the larger the sample size, the larger the number of parameters.

A well-known alternative is the *marginal maximum likelihood* (MML) estimation (Bock & Aitkin, 1981): A distribution $g(\theta)$ for the person parameters is assumed and the resulting situation corresponds to a mixed-effects ANOVA: Item difficulties can be regarded as fixed effects and person abilities as random effects. Thus, IRT models fit into the framework of *generalized linear mixed models* (GLMM) as elaborated in de Boeck and Wilson (2004). By integrating over the ability distribution the random nuisance parameters can be removed from the likelihood equations. This leads to consistent estimates of the item parameters. Further discussions of the MML approach with respect to the CML method will follow.

For the sake of completeness, some other methods for the estimation of the item parameters are the following: Anderson, Li and Vermunt (2007) propose a Pseudo-ML approach, Molenaar (1995) and Linacre (2004) give an overview of various (heuristic) non-ML methods, and Bayesian techniques can be found in Baker and Kim (2004, Chapter 7).

However, back to CML, the main idea behind this approach is the assumption that the raw score r_v is a minimal sufficient statistic for θ_v. Starting from the equivalent multiplicative expression of Equation 1 with $\xi_v = \exp(\theta_v)$ and $\varepsilon_i = \exp(-\beta_i)$, i.e.,

$$P(X_{vi} = 1) = \frac{\xi_v \varepsilon_i}{1 + \xi_v \varepsilon_i}, \tag{7}$$

the following likelihood for the response pattern $\mathbf{x}_v$ for a certain subject v results:

$$P(\mathbf{x}_v \mid \xi_v, \boldsymbol{\varepsilon}) = \prod_{i=1}^{k} \frac{(\xi_v \varepsilon_i)^{x_{vi}}}{1 + \xi_v \varepsilon_i} = \frac{\xi_v^{\,r_v} \prod_{i=1}^{k} \varepsilon_i^{\,x_{vi}}}{\prod_{i=1}^{k} (1 + \xi_v \varepsilon_i)}. \tag{8}$$

Using the notation $\mathbf{y} = (y_1,\ldots,y_k)$ for all possible response patterns with $\sum_{i=1}^{k} y_i = r_v$, the probability for a fixed raw score r_v is

$$P(r_v \mid \xi_v, \boldsymbol{\varepsilon}) = \sum_{\mathbf{y}|r_v} \prod_{i=1}^{k} \frac{(\xi_v \varepsilon_i)^{x_{vi}}}{1+\xi_v \varepsilon_i} = \frac{\xi_v^{\,r_v} \sum_{\mathbf{y}|r_v} \prod_{i=1}^{k} \varepsilon_i^{\,x_{vi}}}{\prod_{i=1}^{k}(1+\xi_v \varepsilon_i)}. \tag{9}$$

The crucial term with respect to numerical solutions of the likelihood equations is the second term in the numerator:

$$\gamma_r(\varepsilon_i) \equiv \sum_{\mathbf{y}|r_v} \prod_{i=1}^{k} \varepsilon_i^{\,x_{vi}} \tag{10}$$

These are the *elementary symmetric functions* (of order r). An overview of efficient computational algorithms and corresponding simulation studies can be found in Liou (1994). The `eRm` package uses the summation algorithm as proposed by Andersen (1972).

Finally, by collecting the different raw scores into the vector **r** the conditional probability of observing response pattern $\mathbf{x}_v$ with given raw score r_v is

$$P(\mathbf{x}_v \mid r_v, \boldsymbol{\varepsilon}) = \frac{P(\mathbf{x}_v \mid \xi_v, \boldsymbol{\varepsilon})}{P(r_v \mid \xi_v, \boldsymbol{\varepsilon})}. \tag{11}$$

By taking the product over the persons (independence assumption), the (conditional) likelihood expression for the whole sample becomes

$$L(\boldsymbol{\varepsilon} \mid \mathbf{r}) = P(\mathbf{x} \mid \mathbf{r}, \boldsymbol{\varepsilon}) = \prod_{v=1}^{n} \frac{\prod_{i=1}^{k} \varepsilon_i^{\,x_{vi}}}{\gamma_{r_v}}. \tag{12}$$

With respect to raw score frequencies n_r and by reintroducing the β-parameters, (12) can be reformulated as

$$L(\boldsymbol{\beta} \mid \boldsymbol{r}) = \frac{\exp\left(\sum_{i=1}^{k} x_{+i}\beta_i\right)}{\prod_{r=0}^{k} \gamma_r^{\,n_r}}, \tag{13}$$

where x_{+i} are the item raw scores. It is obvious that by conditioning the likelihood on the raw scores r, the person parameters completely vanish from the expression. As a consequence, the parameters $\hat{\beta}$ can be estimated without knowledge of the subject's abilities. This issue is referred as *person-free item assessment* and we will discuss this topic within the context of specific objectivity in the next section.

Pertaining to asymptotical issues, it can be shown that under mild regularity conditions (Pfanzagl, 1994) the CML estimates are consistent for $n \to \infty$ and k fixed, unbiased, asymptotically efficient, and normally distributed (Andersen, 1970). For the computation of a Rasch model, comparatively small samples are sufficient to get reliable estimates (Fischer, 1988). Whether the MML estimates are unbiased depends on the correct specification of the ability distribution $g(\theta)$. In case of an incorrect assumption, the estimates are biased which is surely a drawback of this method. If $g(\theta)$ is specified appropriately, the CML and MML estimates are asymptotically equivalent (Pfanzagl, 1994).

Fischer (1981) elaborates on the conditions for the existence and the uniqueness of the CML estimates. The crucial condition for the data matrix is that $\mathbf{X}$ has to be *well-conditioned*. To introduce this issue it is convenient to look at a matrix which is *ill-conditioned*: A matrix is ill-conditioned if there exists a partition of the items into two non-empty subsets such that all of a group of subjects responded correctly to items $i+1,\ldots,k$ ($\mathbf{X}_2$) and all of all other subjects failed for items $1,\ldots,i$ ($\mathbf{X}_3$), i.e.,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \\ \mathbf{X}_3 & \mathbf{X}_4 \end{pmatrix} = \left(\begin{array}{ccc:ccc} & & & 1 & \cdots & 1 \\ & \mathbf{X}_1 & & \vdots & \ddots & \vdots \\ & & & 1 & \cdots & 1 \\ \hdashline 0 & \cdots & 0 & & & \\ \vdots & \ddots & \vdots & & \mathbf{X}_4 & \\ 0 & \cdots & 0 & & & \end{array}\right)$$

Thus, following the definition in Fischer (1981): $\mathbf{X}$ will be called *well-conditioned* if in every possible partition of the items into two nonempty subsets some subjects have given response "1" on some item in the first set and response "0" on some item in the second set. In this case a unique solution for the CML estimates $\hat{\beta}$ exists.

This issue is important for structurally incomplete designs which often occur in practice; different subsets of items are presented to different groups of persons $g = 1,\ldots,G$ where $G \leq n$. As a consequence, the likelihood values have to be computed for each group separately and the joint likelihood is the product over the single group likelihoods. Hence, the likelihood in Equation 13 becomes

$$L(\boldsymbol{\beta} \mid \mathbf{r}) = \prod_{g=1}^{G} \frac{\exp\left(\sum_{i=1}^{k} x_{+i}\beta_i\right)}{\prod_{r=0}^{k} \gamma_{g,r}{}^{n_{g,r}}}. \tag{14}$$

This also implies the necessity to compute the elementary symmetric functions separately for each group. The `eRm` package can handle such structurally incomplete designs.

From the elaborations above it is obvious that from an asymptotical point of view the CML estimates are at least as good as the MML estimates. In the past, computational problems (speed, numerical accuracy) involved in calculating the elementary symmetric functions limited the practical usage of the CML approach (see, e.g., Gustafsson, 1980). Nowadays, these issues are less crucial due to increased computer power.

In some cases MML estimation has advantages not shared by CML: MML leads to finite person parameters even for persons with zero and perfect raw score, and such persons are not removed from the estimation process (Molenaar, 1995). On the other hand the consideration of such persons does not seem meaningful from a substantial point of view since the person parameters are not reliable anymore - for such subjects the test is too difficult or too easy, respectively. Thus, due to these covering effects, a corresponding ability estimation is not feasible. However, if the research goal is to find ability distributions such persons should be regarded and MML can handle this.

When estimates for the person parameters are of interest some care has to be taken if the CML method is used since person parameters cancel from the estimation equations. Usually, they are estimated (once having obtained values for the item parameters) by inserting $\hat{\beta}$ (or equivalently $\hat{\varepsilon}$) into Equation 8 and solving with respect to θ. Alternatively, Bayesian procedures are applicable (Hoijtink, 1995). It is again pointed out that each person in the sample gets an own parameter even though limited by the number of different raw scores. From this perspective, IRT models in general and the Rasch model in particular fit nicely into the framework of *person-oriented research* proposed by Bergman, Magnusson, and El-Khouri (2003).

3.2 CML and specific objectivity

In general, the Rasch model can be regarded as a measurement model: Starting from the (nominally scaled) 0/1-data matrix $\mathbf{X}$, the person raw scores r_v are on an ordinal level. They, in turn, are used to estimate the item parameters $\boldsymbol{\beta}$ which are on an interval scale provided that the Rasch model holds.

Thus, Rasch models allow for comparisons between objects on an interval level. Rasch reasoned on requirements to be fulfilled such that a specific proposition within this context can be regarded as "scientific". His conclusions were that a basic requirement is the "objectivity" of comparisons (Rasch, 1961). This claim contrasts assumptions met in *classical test theory* (CTT). A major advantage of the Rasch model over CTT models is the *sample independence* of the results. The relevant concepts in CTT are based on a linear model for the "true score" leading to some indices, often correlation coefficients, which in turn depend on the observed data. This is a major drawback in CTT. According to Fischer (1974), sample independence in IRT models has the following implications:

- The person-specific results (i.e., essentially θ) do not depend on the assignment of a person to a certain subject group nor on the selected test items from an item pool Ψ.
- Changes in the skills of a person on the latent trait can be determined independently from its base level and independently from the selected item subset $\psi \subset \Psi$.

- From both theoretical and practical perspective the requirement for representativeness of the sample is obsolete in terms of a true random selection process.

Based on these requirements for parameter comparisons, Rasch (1977) introduced the term *specific objectivity*: *objective* because any comparison of a pair of parameters is independent of any other parameters or comparisons; *specifically objective* because the comparison made was relative to some specified frame of reference (Andrich, 1988). In other words, if specific objectivity holds, two persons v and w with corresponding parameters θ_v and θ_w, are comparable independently from the remaining persons in the sample and independently from the presented item subset ψ. In turn, for two items i and j with parameters β_i and β_j, the comparison of these items can be accomplished independently from the remaining items in Ψ and independently from the persons in the sample.

The latter is crucial since it reflects completely what is called sample independence. If we think not only of comparing β_i and β_j but rather to estimate these parameters, we achieve a point where specific objectivity requires a procedure which is able to provide estimates $\hat{\boldsymbol{\beta}}$ that do not depend on the sample. This implies that $\hat{\boldsymbol{\beta}}$ should be computable without the involvement of $\boldsymbol{\theta}$. From Section 3 it becomes clear that CML estimation fulfills this requirement: By conditioning on the sufficient raw score vector $\mathbf{r}$, $\boldsymbol{\theta}$ disappears from the likelihood equation and $L(\boldsymbol{\beta} \mid \mathbf{r})$ can be solved without knowledge of $\boldsymbol{\theta}$. This issue is referred to as *separability of item and person parameters* (see, e.g., Wright & Masters, 1982). Furthermore, separability implies that no specific distribution should be assumed neither for the person nor for the item parameters (Rost, 2000). MML estimation requires such assumptions. At this point it is clear that CML estimation is the only estimation method within the Rasch measurement context fulfilling the requirement of *person-free item calibration* and, thus, it maps the epistemological theory of specific objectivity to a statistical maximum likelihood framework. Note that strictly speaking any statistical result based on sample observations is sample-dependent because any result depends at least on the sample size (Fischer, 1987). The estimation of the item parameters is "sample-independent", a term indicating the fact that the actually obtained sample of a certain population is not of relevance for the statistical inference on these parameters (Kubinger, 1989, p. 23).

3.3 CML and Rasch model tests

Another nice implication of CML estimates is that subsequent test statistics are readily obtained and model tests are easy to carry out. Basically, we have to distinguish between tests on item level and global model tests.

On item level, sample independence reflects the property that by splitting up the sample in, e.g., two parts, the corresponding parameter vectors $\hat{\boldsymbol{\beta}}^{(1)}$ and $\hat{\boldsymbol{\beta}}^{(2)}$ should be the same. Thus, when we want to achieve "Rasch model fit" those items have to be eliminated from the test which differ in the subsamples. This important issue in test calibration can be examined, e.g., by using a graphical model test (Rasch, 1960) which will be illustrated in Section 4.2. Fischer and Scheiblechner (1970) propose a $N(0,1)$-distributed test statistic which compares the item parameters for two subgroups:

$$z = \left(\frac{\beta_i^{(1)} - \beta_i^{(2)}}{\sqrt{Var_i^{(1)} - Var_i^{(2)}}} \right) \tag{15}$$

The variance term in the denominator is based on Fisher's function of "information in the sample". However, as Glas and Verhelst (1995) point out discussing their Wald-type test that this term can be extracted directly from the variance-covariance matrix of the CML estimates. In the same article, other item-based tests can be found and corresponding recent Mantel-Haenszel approaches are proposed in Verguts and de Boeck (2001).

The benefit of subsample invariance can also be taken into consideration when developing model fit statistics. In fact, Andersen (1973) derived his LR-statistic based on the underlying principle of *subgroup homogeneity* in Rasch models: For arbitrary disjoint subgroups $t = 1, \ldots, T$ the parameter estimates $\hat{\boldsymbol{\beta}}_t$ should coincide. Andersen's LR-statistic is given by

$$LR = 2\left(\sum_{t=1}^{T} \log L(\hat{\boldsymbol{\beta}}_t \mid \mathbf{r}, \mathbf{X}_t) - \log L(\hat{\boldsymbol{\beta}} \mid \mathbf{r}, \mathbf{X}) \right). \tag{16}$$

This statistic is asymptotically χ^2-distributed with df equal to the number of parameters estimated in the subgroups minus the number of parameters in the total data set. Another test statistic is proposed by Martin-Löf (1973) which is equivalent to the R_{1c}-statistic by Glas (1988). These model tests are based on the observed and expected frequencies based on the number of subjects belonging to raw score group t ($t = 1, \ldots, T$) and giving a positive response to item i ($i = 1, \ldots, k$). More CML-based tests can be found in Glas and Verhelst (1995) and a recent taxonomy is given in Mair (2006).

Thus, application of CML methods facilitates the testing of specific item fit as well as global model fit. Furthermore, additional item-based hypotheses can be examined in a straightforward manner.

4. The eRm package

4.1 Features of the eRm package

When developing the `eRm` (extended Rasch modeling) package, the aim has been to provide a freely available open source tool to compute CML based Rasch models. A natural choice was the implementation within the statistical analysis and programming environment `R` (R Development Core Team, 2006), an open source implementation of the powerful statistical programming language `S`. The basic program can be downloaded from http://CRAN. R-project.org). Once `R` is installed, the `eRm` package can be downloaded and installed easily from the console, and by typing `help(eRm)` the help files can be examined. An introductory reference to `R` is the book by Venables and Smith (2002).

Embedding `eRm` into the flexible framework of `R` is a crucial benefit over existing standalone programs like WINMIRA (von Davier, 1998), LPCM-WIN (Fischer & Ponocny-Seliger, 1998), and others. The results can be processed towards performing additional customized statistical analyses and computations, results tailored to one's specific needs can be exported into various formats. In addition, R provides a very powerful plot engine (see Murell, 2005) such that the user can easily produce customized plots.

Furthermore, the R system is platform independent, i.e., can be used with diverse operating systems, binary distribution are avalaible for MS Windows (NT, 95 and later), MacOS(X), and various Linuxes. Last but not least, R and all installable packages (about 1000 user contributed and maintained packages, covering most recent statistical technology) on CRAN are completely open source. This implies that users have also full access to the source code.

Another important issue in the development phase was that the package should be flexible enough to allow for CML compatible polytomous generalizations of the basic Rasch model such as the RSM and the PCM. In addition, by introducing a design matrix concept linear extensions of these basic models should be applicable. This approach resulted in including the LLTM, the LRSM and the LPCM as the most general model into the eRm package. For the latter model the CML estimation was implemented which can be used for the remaining models as well (see Section 2). A corresponding graphical representation is given in Figure 2.

An important benefit of the package with respect to linearly extended models is that for certain models the design matrix $\mathbf{W}$ can be generated automatically (LPCM-WIN, Fischer & Ponocny-Seliger, 1998, also allows for specifying design matrices but in case of more complex models this can become a tedious task and the user must have a thorough understanding of establishing proper design structures). For repeated measurement models time contrasts in the eRm can be simply specified by defining the number of measurement points, i.e., mpoints. To regard group contrasts like, e.g., treatment and control groups, a corresponding vector (groupvec) can be specified that denotes which person belongs to which group. However, $\mathbf{W}$ can also be defined by the user.

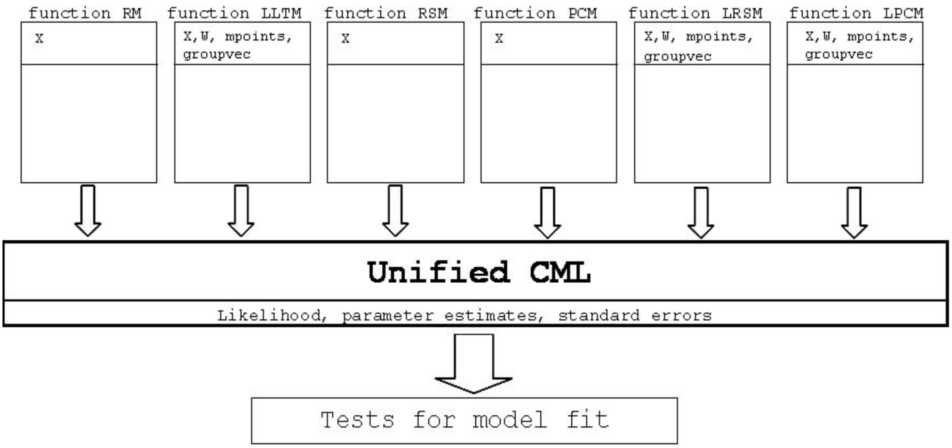

Figure 2: Bodywork of the eRm routine

A recently added feature of the routine is the option to allow for structurally missing values. This is required, e.g., in situations when different subsets of items are presented to different groups of subjects as described in Section 3.1. These person groups are identified automatically: In the data matrix $\mathbf{X}$, those items which are not presented to a certain subject are declared as `NA`s, as usual in `R`.

After solving the CML equations by the Newton-Raphson method, the output of the routine consists of the "basic" parameter estimates $\hat{\boldsymbol{\eta}}$, the corresponding variance-covariance matrix, and consequently the vector with the standard errors. Furthermore, the ordinary item parameter estimates $\hat{\boldsymbol{\beta}}$ are computed by using the linear transformation $\hat{\boldsymbol{\beta}} = \mathbf{W}\hat{\boldsymbol{\eta}}$ For ordinary Rasch models these basic parameters correspond to the item difficulties. It has to be mentioned that the CML equations are solved with the restriction that one item parameter has to be fixed to zero (we use $\beta_1 = 0$). For the sake of interpretability, the resulting estimates $\hat{\boldsymbol{\beta}}$ can easily be transformed into "sum-zero" restricted $\hat{\boldsymbol{\beta}}^*$ by applying $\hat{\beta}_i^* = \hat{\beta}_i - \sum_i \hat{\beta}_i / k$. This transformation is also used for the graphical model test.

In addition, in `eRm` several model tests are implemented as well as a likelihood approach to estimate the person parameters $\hat{\boldsymbol{\theta}}$. As Hoijtink (1995) showed in his simulation studies, Bayesian approaches provide estimates with higher testing power. It is planned to implement such methods in a future version.

4.2 Application examples for dichotomous extended Rasch models

The illustrations here are limited to an ordinary Rasch model and a more parsimonious LLTM on the same data. Finally, it is pointed out how the goodness-of-fit of the LLTM can be evaluated. The corresponding `R` commands are provided. The artificial data matrix $\mathbf{X}$ consists of $n = 15$ persons and $k = 5$ items. It has the following structure:

```
> head (lltmdat)

       [,1] [,2] [,3] [,4] [,5]
[1,]      1    0    0    1    1
[2,]      0    1    1    0    1
[3,]      0    1    0    1    1
[4,]      1    1    1    0    0
[5,]      1    1    0    1    1
[6,]      1    0    1    0    0
          ...
```

For estimating the item parameters of a Rasch model on these data, type the following commands:

```
> resrm <- RM(lltmdat)
> print(resrm)

log-likelihood:  -31.12782

              eta 1     eta 2      eta 3     eta 4
Estimate 0.4944233 0.2446510 -0.2486053 0.2446510
Std.Err  0.7080855 0.7006709  0.7064579 0.7006709
```

To apply Andersen's LR-test the `summary` method can be used as follows.

```
> summary(resrm)

Results of RM fit

number of iterations:  7
log likelihood:  -31.12782
df =  4

Item Parameters:

        Estimate Std. Error    z value  Pr(>|z|)
eta 1  0.4944233  0.7080855  0.6982537 0.4850186
eta 2  0.2446510  0.7006709  0.3491667 0.7269641
eta 3 -0.2486053  0.7064579 -0.3519039 0.7249103
eta 4  0.2446510  0.7006709  0.3491667 0.7269641


Model fit (Andersen test):

LR statistic = 4.046001   df = 4  p = 0.399816
```

Obviously, the Rasch model holds. Coevally, this is a necessary condition for the fit of a more parsimonious LLTM.

The artificial design matrix **W** for an LLTM may be

```
     [,1] [,2]
[1,]    1    2
[2,]    2    2
[3,]    1    1
[4,]    3    1
[5,]    2    1
```

The parameter estimates for the LLTM are

```
> reslltm <- LLTM(lltmdat, lltmdes)
> print(reslltm)

log-likelihood:  -31.65225

              eta 1     eta 2
Estimate -0.09775528 0.1141153
Std.Err   0.31296154 0.4779269
```

To complete the examination of the LLTM model fit, a graphical model test could be performed by plotting the sum-zero restricted parameter estimates of the Rasch model $\hat{\beta}_i^{(RM)}$ against the LLTM estimates $\hat{\beta}_i^{(LLTM)}$

```
> x <- scale(resrm$betapar, scale = FALSE)
> y <- scale (reslltm$betapar, scale = FALSE)
> L <- max(abs(x), abs(y))
> plot(x, y, main = "Graphical LLTM Model Test", xlab =
"Beta RM",
+ ylab = "Beta LLTM", xlim = c(-3, 3), ylim = c(-3, 3),
type = "n")
> text(x, y)
> abline(0, 1)
```

An inspection of the graphical model test (see Figure 3) shows no dramatic deviations, the parameter estimates are not too distant from the diagonal and thus no violation with respect to LLTM homogeneity must be be assumed. Hence, the parameter restrictions imposed in **W** are acceptable.

Further eRm computational examples for polytomous item responses and additional time/group contrasts can be found in Mair and Hatzinger (2007).

5. Discussion

In this paper the open source package eRm implemented in R was introduced. This package uses the CML approach to estimate the item parameters. Benefits of this method were pointed out both from a mathematical and from an epistemological point of view.

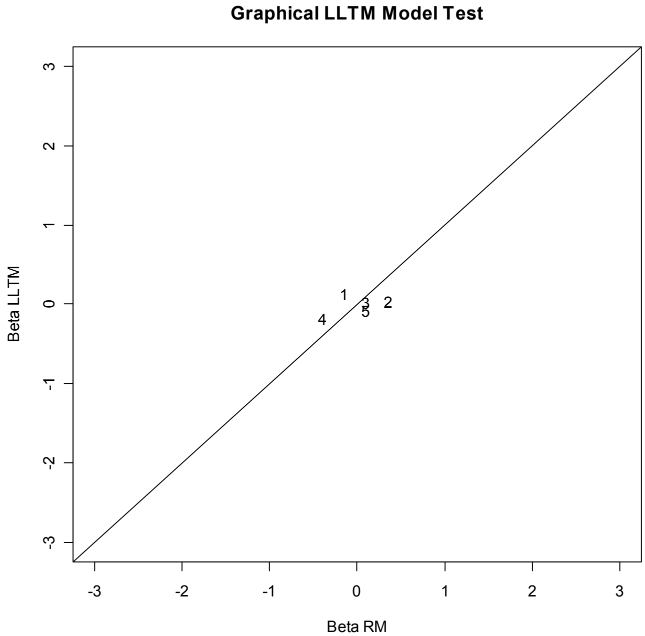

Figure 3: Graphical Model Test

The main purpose of this article was to provide some basics and methodological considerations as our motivation for developing such a package. Therefore, only a fraction of the implemented features was mentioned. A more detailled description can be found in the reference manual (Mair & Hatzinger, 2006). Important recently added features are the treatment of structurally incomplete data and the estimation of person parameters. Amongst other extensions, future versions will implement Bayesian approaches to estimate person parameters (see Hoijtink, 1995) as well as person and item fit indices (Smith, 2004) based on the residuals.

Further models that fit into the `eRm` framework and could be implemented are the following: The *linear logistic model with relaxed assumptions* (Fischer, 1977), abbreviated to LLRA, dispenses the uni-dimensionality requirement of the RM. The reparameterization $\theta_v - \beta_i =: \theta_{vi}$ leads to a generalization of the RM with θ_{vi} as *independent trait parameters*. Applications of this model for the analysis of change as well as the formal equivalence of the LLRA and the LLTM (by introducing the concept if virtual persons) are described in Fischer (1995b). Due to this equivalence, CML estimation can be applied. In combination with the EM-algorithm, the CML approach can also be used to estimate *mixed Rasch models* (MIRA) which emanate from *latent class analysis* (Formann, 1984). The basic idea behind such models is that the extended Rasch model holds within subpopulations of individuals (i.e., latent classes), but with different parameter values for each subgroup (for details see, e.g, Rost & von Davier, 1995). Finally, the implementation of the *one parameter logistic model* (OPLM; Verhelst & Glas, 1995) could be an issue. This is a simplification of the 2-PL model with respect to a fixed item discrimination parameter and consequently, CML estimation is still applicable.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society, Series B, 32, 283-301.

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. Journal of the Royal Statistical Society, Series B, 34, 42-54.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. Psychometrika, 38, 123-140.

Andersen, E. B. (1995). Polytomous Rasch models and their estimation. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 271-292). New York: Springer.

Anderson, C., Li, Z., & Vermunt, J. (2007). Estimation of models in the Rasch family for polytomous items and multiple latent variables. Journal of Statistical Software.

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43, 561-573.

Andrich, D. (1988). Rasch models for measurement. Newbury Park, CA: Sage.

Baker, F. B., & Kim, S. (2004). Item response theory: Parameter estimation techniques (2nd ed.). New York: Dekker.

Bergman, L. R., Magnusson, D., & El-Khouri, B. M. (2003). Studying individual development in an interindividual context: A person-oriented approach. Mahwah, NJ: Lawrence Erlbaum.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord & M. Novick (Eds.), Statistical theories of mental test scores (p. 395-479). Reading, MA: Addison-Wesley.

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

de Boeck, P., & Wilson, M. (2004). Explanatory item response models: A generalized linear and nonlinear approach. New York: Springer.

Embreston, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. Acta Psychologica, 37, 359-374.

Fischer, G. H. (1974). Einführung in die Theorie psychologischer Tests [Introduction to mental test theory]. Bern: Huber.

Fischer, G. H. (1977). Linear logistic trait models: Theory and application. In H. Spada & W. F. Kempf (Eds.), Structural models of thinking and learning (p. 203-225). Bern: Huber.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. Psychometrika, 46, 59-77.

Fischer, G. H. (1987). Applying the principles of specific objectivity and of generalizability to the measurement of change. Psychometrika, 52, 565-587.

Fischer, G. H. (1988). Spezifische Objektivität: Eine wissenschaftstheoretische Grundlage des Rasch-Modells. [Specific objectivity: An epistemological foundation of the Rasch model.]. In K. Kubinger (Ed.), Moderne Testtheorie (p. 87-111). Weinheim: Beltz.

Fischer, G. H. (1995a). Derivations of the Rasch model. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 15-38). New York: Springer.

Fischer, G. H. (1995b). Linear logistic models for change. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundation, recent developments, and applications. New York: Springer.
Fischer, G. H., & Molenaar, I. (1995). Rasch models: Foundations, recent developments, and applications. New York: Springer.
Fischer, G. H., & Parzer P. (1991). An extension of the rating scale model with an application to the measurement of change. Psychometrika, 56, 637-651.
Fischer, G. H. & Ponocny, I. (1994). An extension of the partial credit model with an application to the measurement of change. Psychometrika, 59, 177-192.
Fischer, G. H. & Ponocny-Seliger, E. (1998). Structural Rasch modeling: Handbook of the usage of LPCM-WIN 1.0. Groningen: ProGAMMA.
Fischer, G. H. & Scheiblechner, H. H. (1970). Algorithmen und Programme für das probabilistische Testmodel von Rasch. [Algorithms and programs for Rasch's probabilistic test model.]. Psychologische Beiträge, 12, 23-51.
Formann, A. K. (1984). Die Latent-Class-Analyse. [Latent class analysis.]. Weinheim: Beltz Verlag.
Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. Psychometrika, 53, 525-546.
Glas, C. A. W. & Verhelst. N. (1995). Testing the Rasch model. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 69-96). New York: Springer.
Gustafsson, J. (1980). Testing and obtaining fit of data to the Rasch model. British Journal of Mathematical and Statistical Psychology, 33, 205-233.
Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. The Annals of Statistics, 5, 815-841.
Hoijtink, H. (1995). On person parameter estimation on the dichotomous Rasch model. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 53-68). New York: Springer.
Irtel, H. (1995). An extension of the concept of specific objectivity. Psychometrika, 60, 115-118.
Kubinger, K. D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie. [Current status and critical appreciation of probabilistic test theory]. In K. Kubinger (Ed.), Moderne Testtheorie: Ein Abriss samt neuesten Beiträgen (p. 19-83). Weinheim: Beltz.
Linacre, J. M. (2004). Estimation methods for Rasch measures. In E. V. Smith Jr. & R. M. Smith (Eds.), Introduction to Rasch measurement (p. 25-48). Maple Grove, MN: JAM Press.
Liou, M. (1994). More on the computation of higher-order derivatives of the elementary symmetric functions in the Rasch model. Applied Psychological Measurement, 18, 53-62.
Mair, P. (2006). Simulation studies for goodness-of-fit statistics in item response theory. University of Vienna: Unpublished Master Thesis.
Mair, P., & Hatzinger, R. (2006). eRm: extended Rasch models. R package version 0.3.2: `http://CRAN.R-project.org/`.
Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. Journal of Statistical Software (forthcoming).
Martin-Löf, P. (1973). Statistika modeller. [Statistical models.]. Stockholm: Institutet för Försäkringsmatematik och Mathematisk Statistisk vid Stockhoms Universitet.
Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.
Molenaar, I. (1995). Estimation of item parameters. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 39-51). New York: Springer.
Murell, P. (2005). R graphics. Boca Raton, FL: Chapman & Hall/CRC.

Pfanzagl, J. (1994). On item parameter estimation in certain latent trait models. In G. Fischer & D. Laming (Eds.), Contributions to mathematical psychology, psychometrics, and methodology (p. 249-263). New York: Springer.

R Development Core Team. (2006). R: A language and environment for statistical computing. Vienna, Austria. (ISBN 3-900051-07-0)

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In Proceedings of the IV. Berkeley symposium on mathematical statistics and probability, vol IV (p. 321-333). Berkeley: University of California Press.

Rasch, G. (1977). On specific objectivity: An attempt at formalising the request for generality and validity of scientific statements. Danish Yearbook of Philosophy, 14, 58-94.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests with foreword and afterword by B. D. Wright. Chicago: University of Chicago Press.

Rost, J. (1988). Quantitative und qualitative probabilistische Testtheorie. [Quantitative and qualitative probabilistic test theory.]. Bern: Huber.

Rost, J. (2000). The growing family of Rasch models. In A. Boomsma, M. van Duijin, & T. Snijders (Eds.), Essays on item response theory (p 25-42). New York: Springer.

Rost, J., & von Davier, M. (1995). Polytomous mixed Rasch models. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 371-382). New York: Springer.

Rost, J., & Strauß, B. (1992). Recent developments in psychometrics and test theory. The German Journal of Psychology, 16, 91-119.

Scheiblechner, H (1972). Das Lernen und Lösen komplexer Denkaufgaben. [The learning and solving of complex reasoning items.]. Zeitschrift für Experimentielle und Angewandte Psychologie, 3, 456-506.

Smith, R. M. (2004). Fit analysis in latent trait measurement models. In E. V. Smith Jr. & R. M. Smith (Eds.), Introduction to Rasch measurement (p. 73-92). Maple Grove, MN: Jam Press.

Venables, W. N., & Smith, D. N. (2002). An introduction to R. Bristol: Network Theory Ltd.

Verguts, T., & de Boeck, P. (2001). Some Mantel-Haenszel tests of Rasch model assumptions. British Journal of Mathematics and Statistical Psychology, 54, 21-37.

Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model. In G. Fischer & I. Molenaar (Eds.), Rasch models: Foundations, recent developments, and applications (p. 215-238). New York: Springer.

von Davier, M. (1998). WINMIRA: A Windows program for mixed Rasch models. Kiel: IPN.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis: Rasch measurement. Chicago: Mesa Press.

Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. Educational and Psychological Measurement, 29, 23-48.