Rasch Models
and the R package eRM

Reinhold Hatzinger

Institute for Statistics and Mathematics
WU Vienna
**What is eRm?**

- **eRm** short for *extended Rasch modelling*
- is an *R* package
- is open source: no license fees, source code available, GPL: share, change, and redistribute under certain conditions
- for Rasch family models:
  utilities for fitting, testing, and displaying results
- currently implemented models:
  LPCM, PCM, LRSM, RSM, LLTM, RM, (LLRA)
- uses CML estimation
What is Item Response Theory (IRT)?

IRT is built around the central idea: the probability of a subject’s certain reaction to a stimulus can be described as a function characterising the subject’s location on a latent trait plus one or more parameters characterising the stimulus.
The Rasch Model (RM) (Rasch, 1960)

\[ P(X_{vi} = 1|\theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)} \]

- \( X_{vi} \) ... person \( v \) gives correct answer to item \( i \)
- \( \theta_v \) ... ‘ability’ of person \( v \)
- \( \beta_i \) ... ‘difficulty’ of item \( i \)

<table>
<thead>
<tr>
<th></th>
<th>( I_1 )</th>
<th>( I_2 )</th>
<th>( I_3 )</th>
<th>( I_4 )</th>
<th>( r_v )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_1 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>( P_2 )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>( P_3 )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>( P_4 )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>( s_i )</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>–</td>
</tr>
</tbody>
</table>

Raw Scores:

\[ \sum_i x_{vi} = r_v \]
\[ \sum_v x_{vi} = s_i \]
The Rasch Model

Several ICCs
Rasch Model Assumptions / Properties

**unidimensionality** \( P(X_{vi} = 1|\theta_v, \beta_i, \varphi) = P(X_{vi} = 1|\theta_v, \beta_i) \)
response probability does not depend on other variables \( \varphi \)

**sufficiency** \( f(x_{vi}, ..., x_{vk}|\theta_v) = g(r_v|\theta_v)h(x_{vi}, ..., x_{vk}) \)
raw score \( r_v = \sum_i x_{vi} \) (sum of responses) contains all information on ability, regardless which items have been solved

**conditional independence** \( X_{vi} \perp X_{vj}|\theta_v, \forall i, j \)
for fixed \( \theta \) there is no correlation between any two items

**monotonicity** for \( \theta_v > \theta_w : f(x_{vi}|\theta_v, \beta_i) > f(x_{wi}|\theta_w, \beta_i), \forall \theta_v, \theta_w \)
response probability increases with higher values of \( \theta \)
Parameter Estimation

Item Parameter Estimation

- likelihood based methods:
  differ in their treatment of person parameters
  
  - joint ML estimation (JML)
  - conditional ML estimation (CML)
  - marginal ML estimation (MML)

- other methods available:
  less often used
  not covered here

Person Parameter Estimation

- ML and weighted ML estimation
- Bayes approaches
Joint Maximum Likelihood (JML)

\[ L_u = \frac{\exp(\sum_v \theta_v r_v) \exp(-\sum_i \beta_i s_i)}{\prod_v \prod_i (1 + \exp(\theta_v - \beta_i))} \]

sufficient statistics are:
\[ r_v = \sum_i x_{vi} \text{ for } \theta_v \]
\[ s_i = \sum_v x_{vi} \text{ for } \beta_i \]

problem: item parameter estimates inconsistent as \( n \to \infty \)
biased in finite samples with \( k(k-1) \)

Marginal Maximum Likelihood (MML)

integrate out the person parameter

\[ L_m = \prod_r \left[ \exp(-\sum_i \beta_i s_i) \int \frac{\exp(\theta r)}{\prod_{i=1}^k (1 + \exp(\theta - \beta_i))} dG(\theta) \right]^{n_r} \]

distribution for \( \theta \) must be specified, usually \( \theta \sim N(0,1) \)
can be estimated in R using the ltm package (Rizopoulos, 2009)
Conditional Maximum Likelihood (CML)

condition on $r_v$

$$L_c = \exp(-\sum_i \beta_i s_i) / \prod_r \sum_x \exp(-\sum_i x_i \beta_i)^{n_r}$$

- person parameters do not occur in the conditional likelihood
- items can be compared independent of persons (separation)
- leads to specific objectivity
- person free item calibration
- ‘sample-independence’:
  actual sample not of relevance for inference on item parameters

CML estimates are unbiased and consistent as $n \to \infty$

for estimability set $\beta_1 = 0$ or $\sum \beta_i = 0$

items with score $s_i = 0$ or $n$ and person with $r_v = 0$ or $k$ are removed prior to estimation
MML vs CML

MML Advantages:
- gives also estimates for persons with $r_v = 0$ or $r_v = k$
- when research aims at person distribution
- allows estimation of additional parameters (2PL, 3PL models)
- faster with large $k$

CML Advantages:
- when RM is used as measurement model (scale construction)
- MML parameters can be biased if $G(\theta)$ incorrectly specified
- CML closer to concept of person-free assessment
- allows for specific objectivity
- several goodness-of-fit tests not available with MML

distributional properties of CML and MML estimates are the same asymptotically
Person Parameter Estimation

using the unconditional likelihood

\[ L_u = \frac{\exp(\sum_v \theta_v r_v) \exp(-\sum_i \beta_i s_i)}{\prod_v \prod_i (1 + \exp(\theta_v - \beta_i))} \]

and assuming the \( \beta \)s to be known (from prior estimation)

slightly biased (bias smaller than s.e.’s of estimates)
no estimates for \( r_v = 0 \) and \( r_v = k \)
can be approximated using, e.g., spline interpolation

weighted ML estimation:
likelihood function is skewed, additional source of estimation bias
Warm (1989) suggests unbiasing correction, computationally un-feasible
The R package eRm (extended Rasch modelling)

> library(eRm)

main functions concerning fit of the RM:

- **RM(data)** fits the RM and generates object of class **dRm**
- **person.parameter(drmobj)** generates object of class **ppar**
- plots from **drm** object:
  - `plotPImap()`, `plotICC()`, `plotjointICC()`
- plots from **ppar** object:
  - `plot()`
- extract information from **drm** object:
  - `coef()`, `vcov()`, `confint()`, `logLik()`, `model.matrix()`
- extract information from **ppar** object:
  - `confInt()`, `logLik()`
**Fitting the RM**

> rm.res <- RM(data)
> rm.res

Results of RM estimation:

Call: RM(X = data)

Conditional log-likelihood: -156.3100
Number of iterations: 12
Number of parameters: 4

Basic Parameters eta:

<table>
<thead>
<tr>
<th></th>
<th>eta 1</th>
<th>eta 2</th>
<th>eta 3</th>
<th>eta 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.4292685</td>
<td>-1.1743542</td>
<td>-0.1496732</td>
<td>0.02667262</td>
</tr>
<tr>
<td>Std.Err</td>
<td>0.1945618</td>
<td>0.2243309</td>
<td>0.1918824</td>
<td>0.19118379</td>
</tr>
</tbody>
</table>

– default is: `RM(datamatrix, sum0 = TRUE, other options)`
– `sum0` defines constraints (for estimability):
  - **TRUE** ... sum zero,
  - **FALSE** ... first item set to 0
– the output gives easiness (not difficulty) parameters!
> summary(rm.res)
Results of RM estimation:

Call:  RM(X = data)

Conditional log-likelihood:  -156.3100
Number of iterations:  12
Number of parameters:  4

Basic Parameters (eta) with 0.95 CI:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>lower CI</th>
<th>upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>eta 1</td>
<td>0.429</td>
<td>0.195</td>
<td>0.048</td>
<td>0.811</td>
</tr>
<tr>
<td>eta 2</td>
<td>-1.174</td>
<td>0.224</td>
<td>-1.614</td>
<td>-0.735</td>
</tr>
<tr>
<td>eta 3</td>
<td>-0.150</td>
<td>0.192</td>
<td>-0.526</td>
<td>0.226</td>
</tr>
<tr>
<td>eta 4</td>
<td>0.027</td>
<td>0.191</td>
<td>-0.348</td>
<td>0.401</td>
</tr>
</tbody>
</table>

Item Easiness Parameters (beta) with 0.95 CI:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>lower CI</th>
<th>upper CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta I1</td>
<td>0.868</td>
<td>0.206</td>
<td>0.464</td>
<td>1.272</td>
</tr>
<tr>
<td>beta I2</td>
<td>0.429</td>
<td>0.195</td>
<td>0.048</td>
<td>0.811</td>
</tr>
<tr>
<td>beta I3</td>
<td>-1.174</td>
<td>0.224</td>
<td>-1.614</td>
<td>-0.735</td>
</tr>
<tr>
<td>beta I4</td>
<td>-0.150</td>
<td>0.192</td>
<td>-0.526</td>
<td>0.226</td>
</tr>
<tr>
<td>beta I5</td>
<td>0.027</td>
<td>0.191</td>
<td>-0.348</td>
<td>0.401</td>
</tr>
</tbody>
</table>
Extracting Information

the item parameter estimates

```r
> coef(rm.res)
    eta 1     eta 2     eta 3     eta 4
 0.42926853 -1.17435425 -0.14967319  0.02667262
```

the variance-covariance matrix of item parameter estimates

```r
> vcov(rm.res)
[1,] 0.037854306 -0.01255417 -0.008073628 -0.007959444
[2,] -0.012554175 0.05032436 -0.011716057 -0.011780088
[3,] -0.008073628 -0.01171606  0.036818867 -0.007484464
[4,] -0.007959444 -0.01178009 -0.007484464  0.036551241
```
Extracting Information (cont’d)

confidence intervals for the item parameter estimates

> confint(rm.res, "beta")

<table>
<thead>
<tr>
<th>beta</th>
<th>2.5 %</th>
<th>97.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>0.4644</td>
<td>1.2717</td>
</tr>
<tr>
<td>I2</td>
<td>0.0479</td>
<td>0.8106</td>
</tr>
<tr>
<td>I3</td>
<td>-1.6140</td>
<td>-0.7347</td>
</tr>
<tr>
<td>I4</td>
<td>-0.5258</td>
<td>0.2264</td>
</tr>
<tr>
<td>I5</td>
<td>-0.3480</td>
<td>0.4014</td>
</tr>
</tbody>
</table>

the conditional log likelihood

> logLik(rm.res)

'Conditional log Lik.' -156.3100 (df=4)
Plot ICCs

```r
> plotjointI(rm.res, xlim = c(-5, 5))
```
Plot single ICC

> plotI(rm.res, i = 3)
Plot ICCs

> plotICC(rm.res, item.subset = 1:4, ask = F, empICC = list("raw"),
+   empCI = list(lty = "solid"))
Plot Person-Item Map

> plotPImap(rm.res)
Person Parameter Estimation

```r
> pp <- person.parameter(rm.res)
> pp
Person Parameters:

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Estimate</th>
<th>Std.Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2.6310979</td>
<td>NA</td>
</tr>
<tr>
<td>1</td>
<td>-1.5189967</td>
<td>1.1498599</td>
</tr>
<tr>
<td>2</td>
<td>-0.4615091</td>
<td>0.9565426</td>
</tr>
<tr>
<td>3</td>
<td>0.4374933</td>
<td>0.9636302</td>
</tr>
<tr>
<td>4</td>
<td>1.5217580</td>
<td>1.1669396</td>
</tr>
<tr>
<td>5</td>
<td>2.6659917</td>
<td>NA</td>
</tr>
</tbody>
</table>
```

if NAs in the data, different person parameters are estimated for every NA-pattern group
Methods for Person Parameter Estimation Results

> logLik(pp)
'Unconditional (joint) log Lik.' -10.85398 (df=4)

> confint(pp)


2.5 % 97.5 %
P1 -3.772681 0.7346872
P2 -1.451187 2.3261739
P3 -1.451187 2.3261739
P5 -2.336298 1.4132799
P6 -1.451187 2.3261739
P7 -2.336298 1.4132799
...

attention: confint(pp) gives values for all subjects if there are Nas in the data, confidence intervals are printed for each NA group
Plot of Person Parameter Estimates

> plot(pp)
Testing the RM – Overview

RM allows to evaluate the quality of measurement

crucial assumptions empirically testable

aim: find set of items that conform to the RM (‘data fit model’)

various tests/diagnostics have been proposed

some implemented in eRm:

– Andersen LR test
– Wald-type test
– nonparametric tests
– item/person fit indices
– graphical procedures

Munich 2010
Andersen’s Likelihood Ratio Test (Andersen, 1973)

– ‘global’ test (all items investigated simultaneously)
– powerful against violations of sufficiency and monotonicity
– can detect DIF (differential item functioning or item bias):

basic idea:
consistent item parameter estimates (‘invariance’) obtained from any subgroup where the model holds

divide the sample according to score \( r, \ r = 1, \ldots, J - 1 \)
obtain \( J - 1 \) likelihoods of the form

\[
L_c^{(r)} = \exp(- \sum_j \beta_j s_j^{(r)})/\gamma(r; \beta_1, \ldots, \beta_J)^{nr}
\]

the total likelihood is \( L_c = \prod_r L_c^{(r)} \)
Andersen’s Likelihood Ratio Test (cont’d)

\[ \Lambda = \frac{L_c}{\prod_r L_c^{(r)}} = 1, \text{ only if the RM holds} \]

\[ Z = -2\ln \Lambda \text{ is asymptotically } \chi^2 \text{-distributed with } df = (J - 2)(J - 1) \]

test can be used for any partition of the sample according to extraneous variables (e.g., gender, age, ...)

Wald Test

allows for testing single items idea is again: sample into subgroups (usually 2)

using separate estimates \( \hat{\beta}_j^{(1)} \) and \( \hat{\beta}_j^{(2)} \) (and \( \hat{\sigma}_{\beta_j}^{(1)} \), \( \hat{\sigma}_{\beta_j}^{(2)} \)),

\[ S_j = (\hat{\beta}_j^{(1)} - \hat{\beta}_j^{(2)})/\sqrt{\hat{\sigma}_{\beta_j}^{(1)} + \hat{\sigma}_{\beta_j}^{(2)}} \approx N(0, 1) \]
Nonparametric (‘exact’) Tests

Idea:
- Parameter estimates depend only on marginals $r$ and $s$
- for any statistic of the data matrix, one can approximate the null distribution
- take random sample from the collection of equally likely data matrices, compute null distribution of statistic
- valid and powerful, even in small samples

Person/Item Fit

objective is to detect noticeable patterns

Expected response: $\pi_{vi} = \exp(\theta_v - \beta_i)/(1 + \exp(\theta_v - \beta_i))$

Residuals: $e_{vi} = x_{vi} - \pi_{vi}$

Example: Outfit MSQ for items: $u_i = \frac{1}{n} \sum_v \frac{e_{vi}^2}{\pi_{vi}(1 - \pi_{vi})}$

test statistics, e.g., $nu_i^2$, are $\chi^2$ with corresponding $df$
Graphical Procedure

underlying idea again subgroup homogeneity, plot $\hat{\beta}^{(1)}$ vs $\hat{\beta}^{(2)}$
**Likelihoodratio- and Wald Tests**

**LR Test:**

```r
> lrt <- LRtest(rm.res, se = TRUE)
> lrt

Andersen LR-test:
LR-value: 2.407
Chi-square df: 4
p-value: 0.661
```

**Wald Test:**

```r
> Waldtest(rm.res)
Wald test on item level (z-values):

<table>
<thead>
<tr>
<th>z-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta I1</td>
<td>-0.832</td>
</tr>
<tr>
<td>beta I2</td>
<td>-0.352</td>
</tr>
<tr>
<td>beta I3</td>
<td>0.428</td>
</tr>
<tr>
<td>beta I4</td>
<td>1.300</td>
</tr>
<tr>
<td>beta I5</td>
<td>-0.411</td>
</tr>
</tbody>
</table>
```

Munich 2010
Item Fit Statistics

> itemfit(pp)
Itemfit Statistics:

<table>
<thead>
<tr>
<th></th>
<th>Chisq</th>
<th>df</th>
<th>p-value</th>
<th>Outfit MSQ</th>
<th>Infit MSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>80.94</td>
<td>84</td>
<td>0.574</td>
<td>0.952</td>
<td>0.966</td>
</tr>
<tr>
<td>I2</td>
<td>78.49</td>
<td>84</td>
<td>0.649</td>
<td>0.923</td>
<td>0.934</td>
</tr>
<tr>
<td>I3</td>
<td>82.48</td>
<td>84</td>
<td>0.526</td>
<td>0.970</td>
<td>0.961</td>
</tr>
<tr>
<td>I4</td>
<td>85.14</td>
<td>84</td>
<td>0.445</td>
<td>1.002</td>
<td>1.024</td>
</tr>
<tr>
<td>I5</td>
<td>74.28</td>
<td>84</td>
<td>0.767</td>
<td>0.874</td>
<td>0.908</td>
</tr>
</tbody>
</table>

Nonparametric Tests

> t11 <- NPtest(data, method = "T11")
> t11
Nonparametric RM model test: T11 (global test - local dependence)
  (sum of deviations between observed and expected inter-item correlations)

Number of sampled matrices: 500
one-sided p-value: 0.934
**Graphical Procedure**

```r
> plotGOF(lrt, conf = list())
```

![Graphical Model Check](image-url)
Polytomous Models

Partial Credit Modell (PCM)

\[ P(X_{vi} = h) = \frac{\exp[h(\theta_v + \beta_i) + \omega_{hi}]}{\sum_{l=0}^{m_i} \exp[l(\theta_v + \beta_i) + \omega_{li}]} \]

- \( h \) ... response categories (\( h = 0, \ldots, m_i \))
- \( m_i \) ... number of response categories may differ across items
- \( \omega_{hi} \) ... category parameter

Rating Scale Modell (RSM)

simplification:

- \( m_i = m \) ... distances between categories are equal across all item
- \( \omega_{hi} = \omega_h \) ... ‘equistant scoring’
ICCs for the PCM

ICC plot for item I2

Latent Dimension

probability for responding in category

Category 0
Category 1
Category 2
Category 3
Comparison RSM vs PCM

RSM

PCM

Munich 2010
**R commands**

main functions concerning fit of polytomous models:

- `PCM(data)` fits the PCM and generates object of class `Rm`
- `RSM(data)` fits the RSM and generates object of class `Rm`
- `thresholds(rmobj)` displays the itemparameter estimates as thresholds
- all other functions are the same as previously presented (except for `plotjointICC()`)

Munich 2010
**eRm Summary**

Core of eRm is the **Linear Partial Credit Model (LPCM):**

\[
P(X_{vi} = h) = \frac{\exp(h\theta_v + \beta_{ih})}{\sum_{l=0}^{m_i} \exp(l\theta_v + \beta_{ih})}
\]

where the $\beta_{ih}$'s are linearly reparameterised

\[
\beta_{ih} = \sum_{p} w_{ihp} \eta_p
\]

allows for a general algorithm:

according to the specification of the design matrix $W = ((w_{ih,p}))$

various models can be estimated

currently functions for:

LPCM, PCM, LRSM, RSM, LLTM, RM
The model hierarchy in eRm

the LPCM is the most general unidimensional model in this family all other models are submodels they are obtained by appropriately defining the design matrix $W$
eRm Features

Scope:
- Scale Analysis (measurement models)
- Modelling latent change (statistical models)
  uni- and multidimensional (LLRA)

Models:
- RM, RSM, PCM, LLTM, LRSM, LPCM, (LLRA)
- Treatment of missing values (MCAR)
- Different constraints for parameter estimation
- Design matrix (default / user defined)

Estimation:
- Itemparameters, ‘basic’- and effect parameters,
  threshold parameters (all using CML)
- Personparameters (JML)
- Covariance matrices (confidence intervals)
- Support for stepwise item selection
eRm Features (cont’d)

Diagnostics, Model Tests, and Fit Statistics:
- Andersen LR-test, Wald Test for single items
- Global and item level nonparametric tests (for RM)
- Itemfit, Personfit (using Pearson residuals)
- Information criteria (AIC, BIC, cAIC)
- Check for existence of ML estimates – ‘well-conditioned datamatrix’ (for RM)
- some (nonpsychometric) logistic regression diagnostics

Plots:
- Goodness-of-Fit Plots
- ICC-Plots for single items (with optional empirical ICCs)
- Joint ICC-Plot (for RM)
- Person-Item Map

Miscellaneous:
- Simulation of data matrices according to RM violations
- ...
Further Infos:

R Forge:  http://r-forge.r-project.org/
  Development platform
  latest releases downloads
  Discussion and help forum
  Project homepage http://erm.r-forge.r-project.org/

Publications:
  Mair & and Hatzinger (2007). Psychology Science