Shrinkage estimation of the three-parameter logistic model

Michela Battauz (joint with Ruggero Bellio)

Department of Economics and Statistics
University of Udine (Italy)

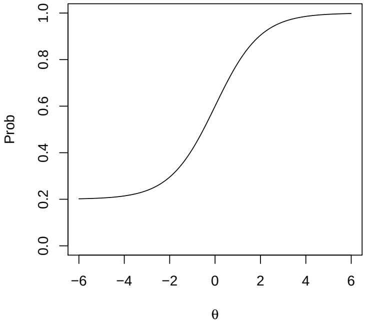
Psychoco 2020

The three-parameter logistic (3PL) model

- Used for modelling the responses of a **proficiency test** with **binary response** items, when the probability of **guessing** is not zero.
- Probability of a correct response $p_{ij} = \Pr(X_{ij} = 1 | \theta_i; a_j, b_j, c_j)$ to item j

$$p_{ij} = c_j + (1 - c_j)\frac{\exp\{a_j(\theta_i - b_j)\}}{1 + \exp\{a_j(\theta_i - b_j)\}},$$

- θ_i ability of person i
- Item parameters:
  - a_j discrimination parameter
  - b_j difficulty parameter
  - c_j guessing parameter



Maximum likelihood estimation

- A convenient **parameterization** of the model, suitable for estimation is

$$p_{ij} = c_j + (1 - c_j)\frac{\exp(\beta_{1j} + \beta_{2j}\theta_i)}{1 + \exp(\beta_{1j} + \beta_{2j}\theta_i)},$$

with

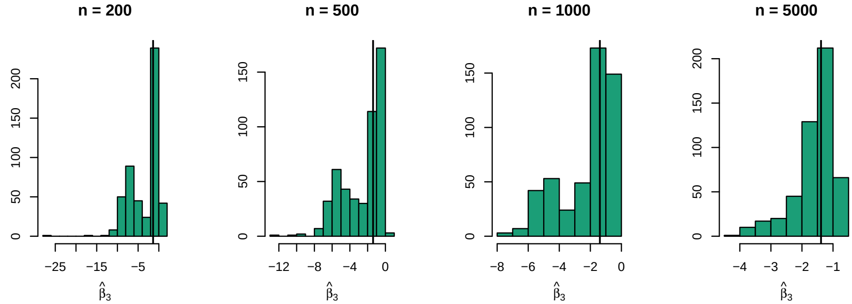
$$c_j = \frac{\exp(\beta_{3j})}{1 + \exp(\beta_{3j})}.$$

- **Marginal Maximum Likelihood** Estimation (**MLE**) (Bock and Aitkin, 1981) usally adopted, where $\theta_i \sim N(0,1)$ and the log-likelihood function is

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log \int_{\mathbb{R}} \prod_{j=1}^{J} p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} \phi(\theta_i) d\theta_i,$$

The 3PL model in practice

- Broadly used in applications
- Included in all major IRT software, including R packages `ltm`, `mirt` and `TAM`.
- Guessing parameters **weakly identifiable** (Patz and Junker, 1999), MLE has convergence problems.
- The estimates of the guessing parameters tend to have a **negative bias**.



Our proposal: shrinkage estimation of the 3PL model

Two main approaches:

- **Penalized maximum likelihood estimation**
- **Model-based shrinkage estimation**

Penalty on the guessing parameters

- Penalized log-likelihood: $\ell_p(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) + J(\boldsymbol{\beta}_3)$
- Penalty proportional to the log p.d.f. of the normal distribution

$$J(\boldsymbol{\beta}_3) = -\frac{1}{2\,\sigma^2}\sum_{j=1}^{J}(\beta_{3j} - \mu)^2\,,$$

(implemented also in the `mirt` package)

- Ridge-type penalty

$$J(\boldsymbol{\beta}_3) = -\lambda\sum_{j<k}(\beta_{3j} - \beta_{3k})^2\,,$$

- The two penalties are equivalent when $\mu = J^{-1}\sum_j \beta_{3j}$, but the ridge-type penalty has only one tuning parameter.
- Empirical results were very similar, so we **chose the ridge-type penalty**.

Model-based shrinkage estimation

- Application of the **bias-reduction** method (**BR**) proposed by Firth (1993) for a general parametric model, with estimating equation for $\boldsymbol{\beta}$

$$\mathbf{S}(\boldsymbol{\beta}) = \nabla \ell(\boldsymbol{\beta}) - \mathcal{I}(\boldsymbol{\beta})^{-1} b(\boldsymbol{\beta}),$$

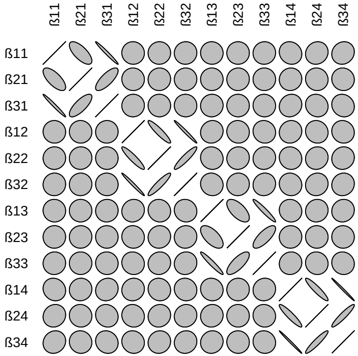
  where $\mathcal{I}(\boldsymbol{\beta})$ is the expected Fisher information, and $b(\boldsymbol{\beta})$ is the leading term of the asymptotic bias of the MLE.
- The estimator defined by solving $\mathbf{S}(\boldsymbol{\beta}) = 0$ has reduced finite sample bias, though it is asymptotically equivalent to the MLE.
- In many models for discrete data, a useful side effect of bias reduction is **shrinkage** of parameter estimates (Kosmidis, 2014).

Implementation

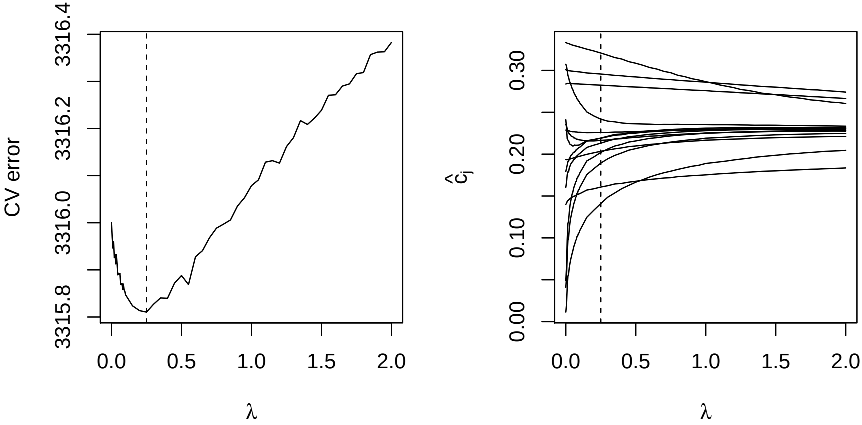
- Both approaches were implemented the **R package S3PL** `github.com/micbtz/S3PL`.
- Integral approximated using **Gaussian quadrature**.
- Penalized likelihood:
  - package **Rcpp** to speed up computational time;
  - tuning parameter λ selected using **cross-validation**;
  - **cross-validation error**: the negative log-likelihood
- Bias reduction:
  - package **TMB** for automatic differentiation and C++ implementation;
  - **Monte Carlo** evaluation of required expected values;
  - **Parallel** computing;

An illustrative example

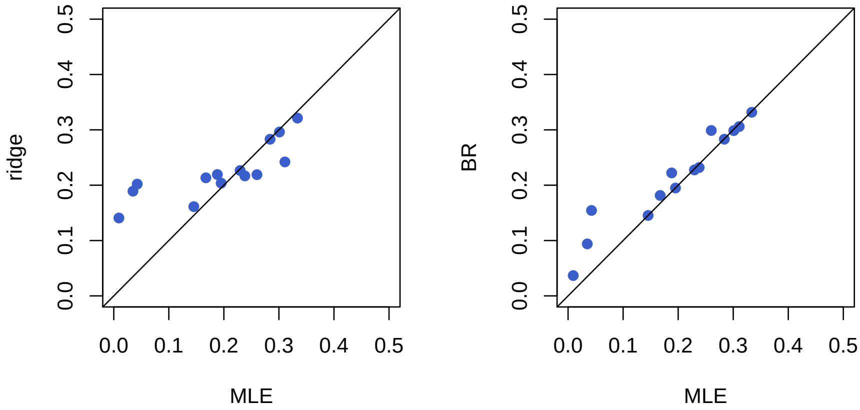
- Achievement data collected on students attending the third year of high school in Italy, tested in Mathematics
- $n = 3843$ students, $J = 14$ items
- Estimated correlation matrix of $\hat{\beta}_{MLE}$



Ridge-type penalization



Comparison of estimates of the guessing parameters



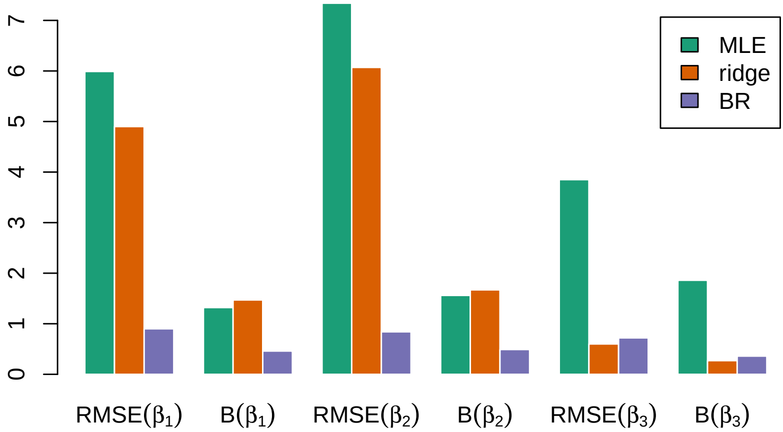
- Ridge-type penalization yields a larger shrinkage of the parameters than BR.

A simulation study

- True item parameters of 30 items taken from TIMSS 2015, Fourth Grade, Mathematics
- True guessing parameters:
  - TIMSS (very variable)
  - constant $c_j = 0.2 \ \forall j$
- Sample size $n = 200, 500, 1000$
- 500 replications for each setting

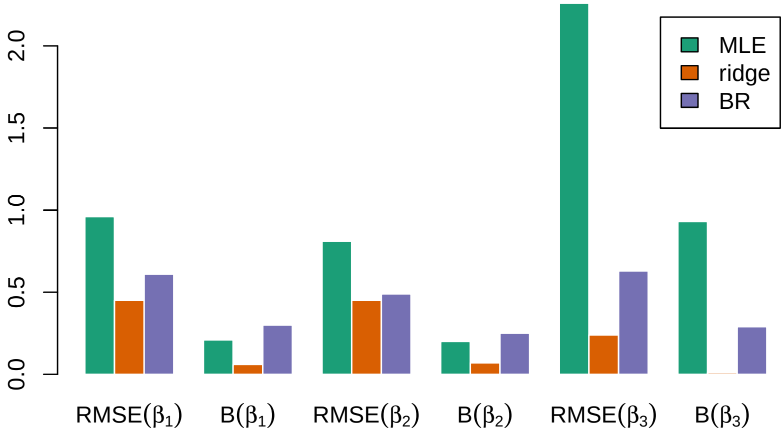
Results of the simulation study

- $n = 200$
- constant guessing parameters



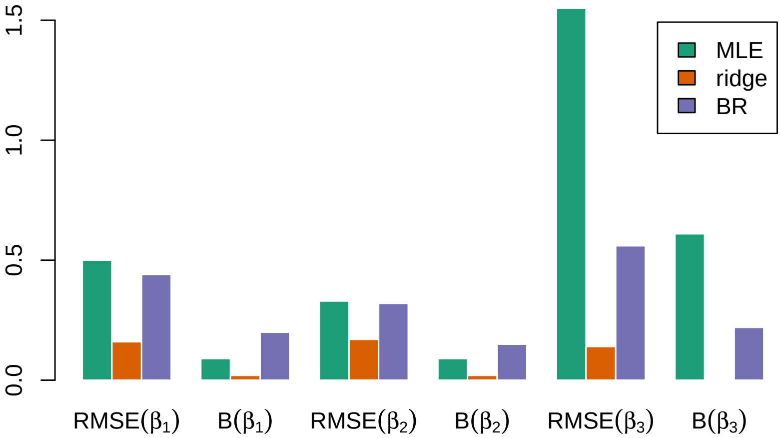
Results of the simulation study

- $n = 500$
- constant guessing parameters



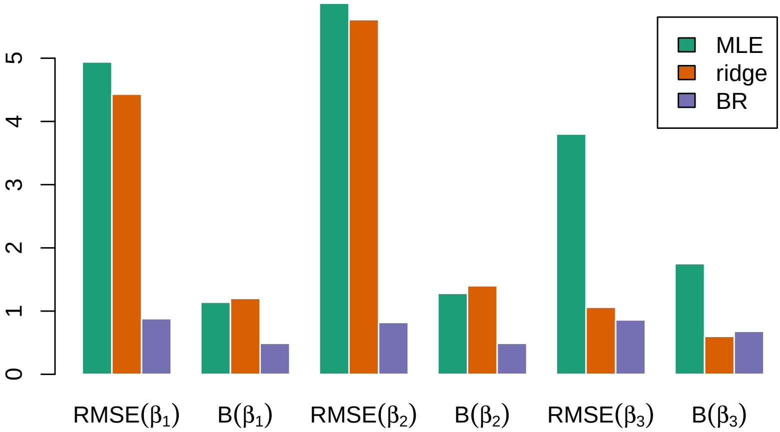
Results of the simulation study

- $n = 1000$
- constant guessing parameters



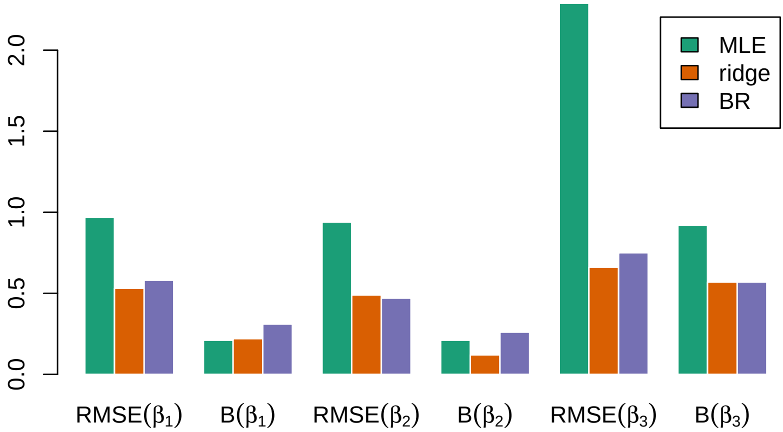
Results of the simulation study

- $n = 200$
- variable guessing parameters



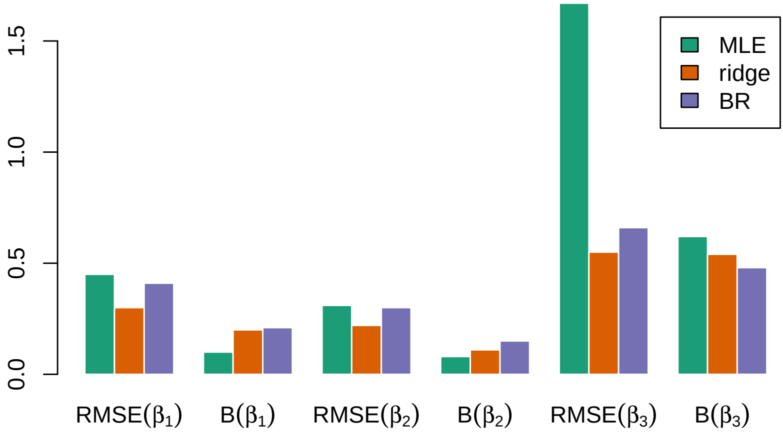
Results of the simulation study

- $n = 500$
- variable guessing parameters



Results of the simulation study

- $n = 1000$
- variable guessing parameters



Conclusions

- MLE seems inaccurate even for large sample sizes.
- The BR method performs well for small sample sizes.
- For larger samples, ridge-type penalty performs better, especially when the true guessing parameters are constant.

References

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

Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80, 27-38.

Kosmidis, I. (2014). Bias in parametric estimation: reduction and useful side-effects. Wires Comp. Statist., 6, 185-196.

Patz, R. J., Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. J. Educ. Behav. Stat., 24, 342–366.

Thank you for your attention!