Maximum penalised likelihood in Factor Analysis

Philipp Sterzinger
Department of Statistics, LSE

p.sterzinger@lse.ac.uk
psterzinger.at psterzinger

IWSM 2026
Oslo, Norway

29 June 2026

Joint with

Irini Moustaki
LSE

Ioannis Kosmidis
Warwick

This talk

Sterzinger P., Kosmidis I., and Moustaki I. (2026). *Maximum Softly Penalized Likelihood in Factor Analysis*, Psychometrika. DOI:10.1017/psy.2026.10092

Normal-linear factor model

Normal-linear factor model

Data

$x_1, x_2, \ldots, x_n \in \mathfrak{R}^p$

Model

$$X_i = \mu_0 + \Lambda_0 Y_i + E_i$$

Unobservables

Latent factors: $Y_i \sim \mathrm{N}(0_q, I_q)$

Unique errors: $E_i \sim \mathrm{N}(0_p, \Psi_0)$

Parameters

Mean vector: $\mu_0 \in \mathfrak{R}^p$

Factor loadings: $\Lambda_0 \in \mathfrak{R}^{p \times q}$

Specific variances Ψ_0: $p \times p$ diagonal, positive definite matrix

Normal-linear factor model

Log-likelihood

$X_i \sim \mathrm{N}(\mu_0, \Sigma_0)$, for $\Sigma_0 = \Lambda_0 \Lambda_0{}^\top + \Psi_0$ so that

$$\ell(\Sigma; S) = -\frac{n}{2}\left\{\log\det(\Sigma) + \operatorname{tr}\left(\Sigma^{-1} S\right)\right\} + C\,,$$

where $S = 1/n \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top$, $\bar{x} = 1/n \sum_{i=1}^{n} x_i$

Maximum likelihood (ML) estimator

$\hat{\Lambda}, \hat{\Psi} = \arg\max \ell(\Lambda\Lambda^\top + \Psi; S)$

for $\Lambda \in \mathfrak{R}^{p \times q}$, $\Psi \in \mathfrak{R}^{p \times p}$ diagonal and positive definite

Motivating Example

Job application data

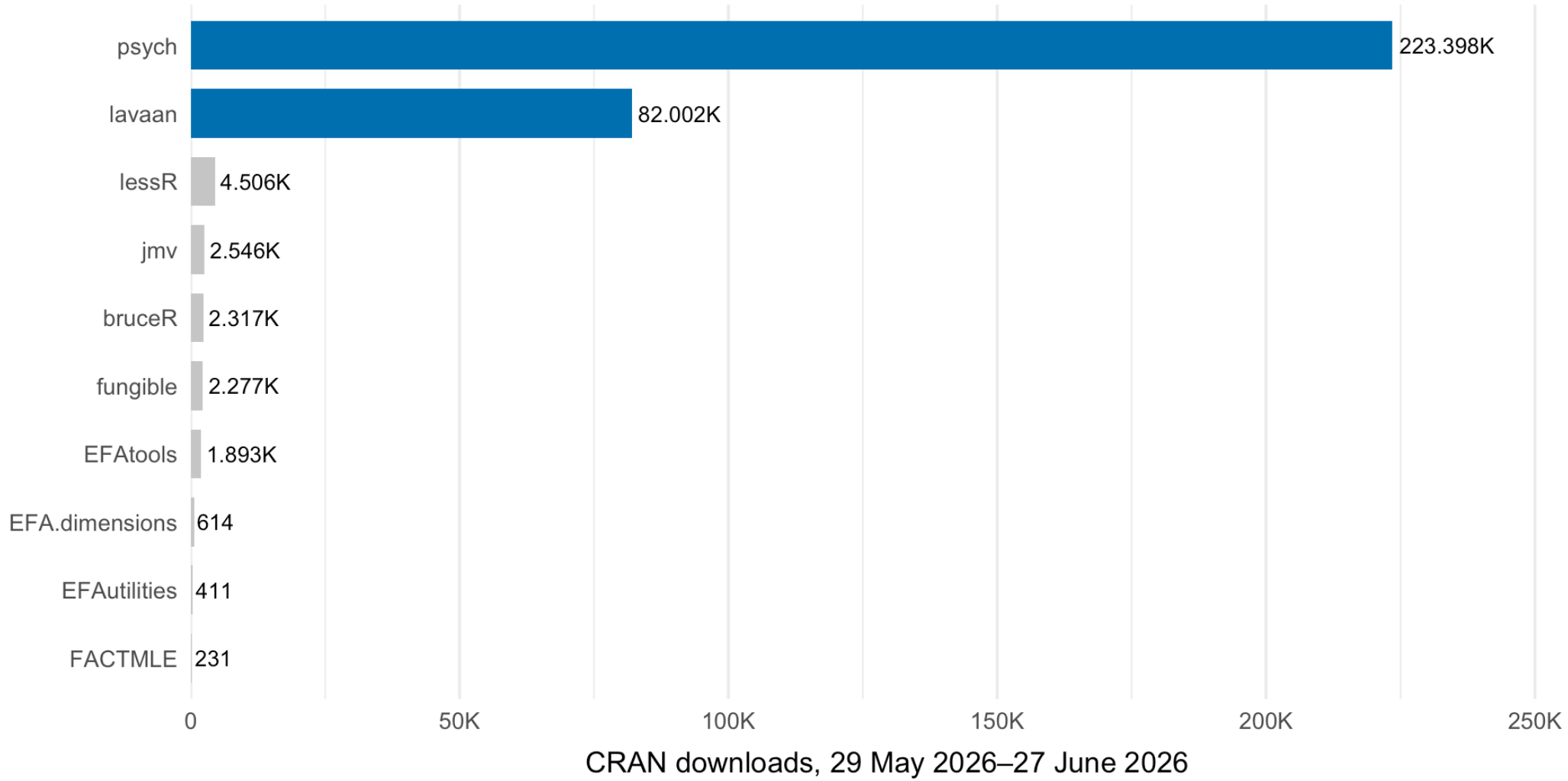
Data from Kendall (1980) which scores job applicants $(n = 48)$ according to a range of criteria $(p = 15)$

Form	Appearance	Ability	Likeability	Selfconf.	Lucidity	Honesty	Salesmanshi	Experience	Drive	Ambition	Grasp	Potential	Keenness	Suitability
6	7	2	5	8	7	8	8	3	8	9	7	5	7	10
9	10	5	8	10	9	9	10	5	9	9	8	8	8	10
7	8	3	6	9	8	9	7	4	9	9	8	6	8	10

Want to use exploratory factor analysis to find common factors underlying the data

Exploratory Factor Analysis with R

Most popular R packages for ML estimation of EFA on CRAN*:



* by monthly downloads; CRAN metadata search for EFA keywords & EFA fit function in documentation; `stats::factanal` excluded

Exploratory Factor Analysis with R

stats::factanal

```
factanal_fit <- stats::factanal(x = job_data, factors = 4)
```

```
factanal_uniquenesses
```

```
      Form Appearance    Ability Likeability  Selfconf.
0.50334512 0.70264896 0.57633097 0.27770086 0.12490328
   Lucidity    Honesty Salesmanship Experience      Drive
0.17331439 0.68782086 0.10727743 0.37295753 0.21792774
   Ambition      Grasp  Potential   Keenness Suitability
0.15640507 0.13156330 0.09499911 0.00500000 0.14549804
```

Exploratory Factor Analysis with R

lavaan::efa

```
lavaan_fit <- efa(data = job_data, nfactors = 4)
```

```
lavaan_uniquenesses
```

```
        Form  Appearance      Ability Likeability    Selfconf.
   3.5220432    2.6588417    2.4236585    2.1569829    0.7153340
     Lucidity       Honesty Salesmanship   Experience        Drive
   1.6915113  150.1761574    1.2578323    4.0055043    1.8532189
    Ambition        Grasp    Potential     Keenness  Suitability
   1.3201438    1.1856554    0.9438065    0.0000000    1.5426059
```

Exploratory Factor Analysis with R

psych::fa

```
1  fa_fit <- fa(job_data, nfactors = 4)
```

```
Warning: An ultra-Heywood case was detected. Examine the results carefully
```

```
1  fa_fit$uniquenesses
```

```
      Form Appearance    Ability Likeability   Selfconf.
0.50932136 0.70658334 0.57678509 0.29729996 0.13326597
   Lucidity    Honesty Salesmanship Experience      Drive
0.19702315 0.64029962 0.09690817 0.25028997 0.24868307
   Ambition      Grasp  Potential    Keenness Suitability
0.15787323 0.13769861 0.09571172 -0.00448535 0.20726449
```

Exploratory Factor Analysis with R

psych::fa

```
fa_fit <- fa(job_data, nfactors = 4)
```

```
Warning: An ultra-Heywood case was detected. Examine the results carefully
```

```
fa_fit$uniquenesses
```

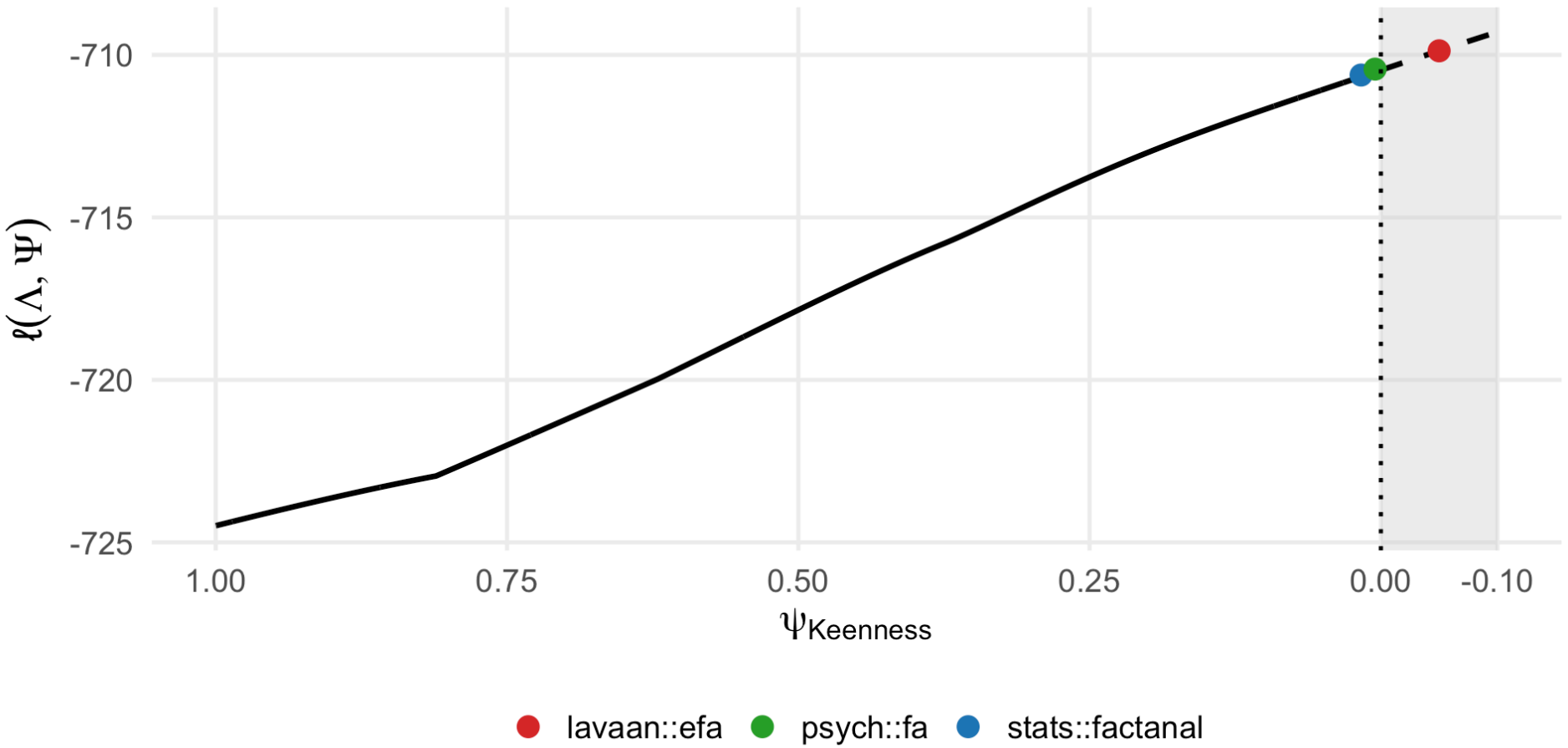
```
        Form   Appearance      Ability Likeability    Selfconf. 
 0.509321362  0.706583340  0.576785093  0.297299960  0.133265973 
   Lucidity      Honesty Salesmanship   Experience        Drive 
 0.197023150  0.640299618  0.096908170  0.250289966  0.248683074 
   Ambition        Grasp    Potential     Keenness  Suitability 
 0.157873233  0.137698609  0.095711717 -0.004485347  0.207264486 
```

The ML estimate does not exist

This phenomenon is termed a **Heywood (1931) case**

Nonexistence of ML estimates

Dominating parameter sequence



Profiled boundary path $(\Lambda_\psi, \Psi_\psi)$ from maximising $\ell(\Lambda, \Psi)$ at fixed $\psi_{\text{Keenness}} = \psi$

Maximum softly-penalized likelihood

Maximum penalized likelihood

Add penalty function $P(\Lambda, \Psi)$ that diverges to $-\infty$ whenever $\Psi \to 0$, or $\Psi, \Lambda \to \infty$

Penalized log-likelihood

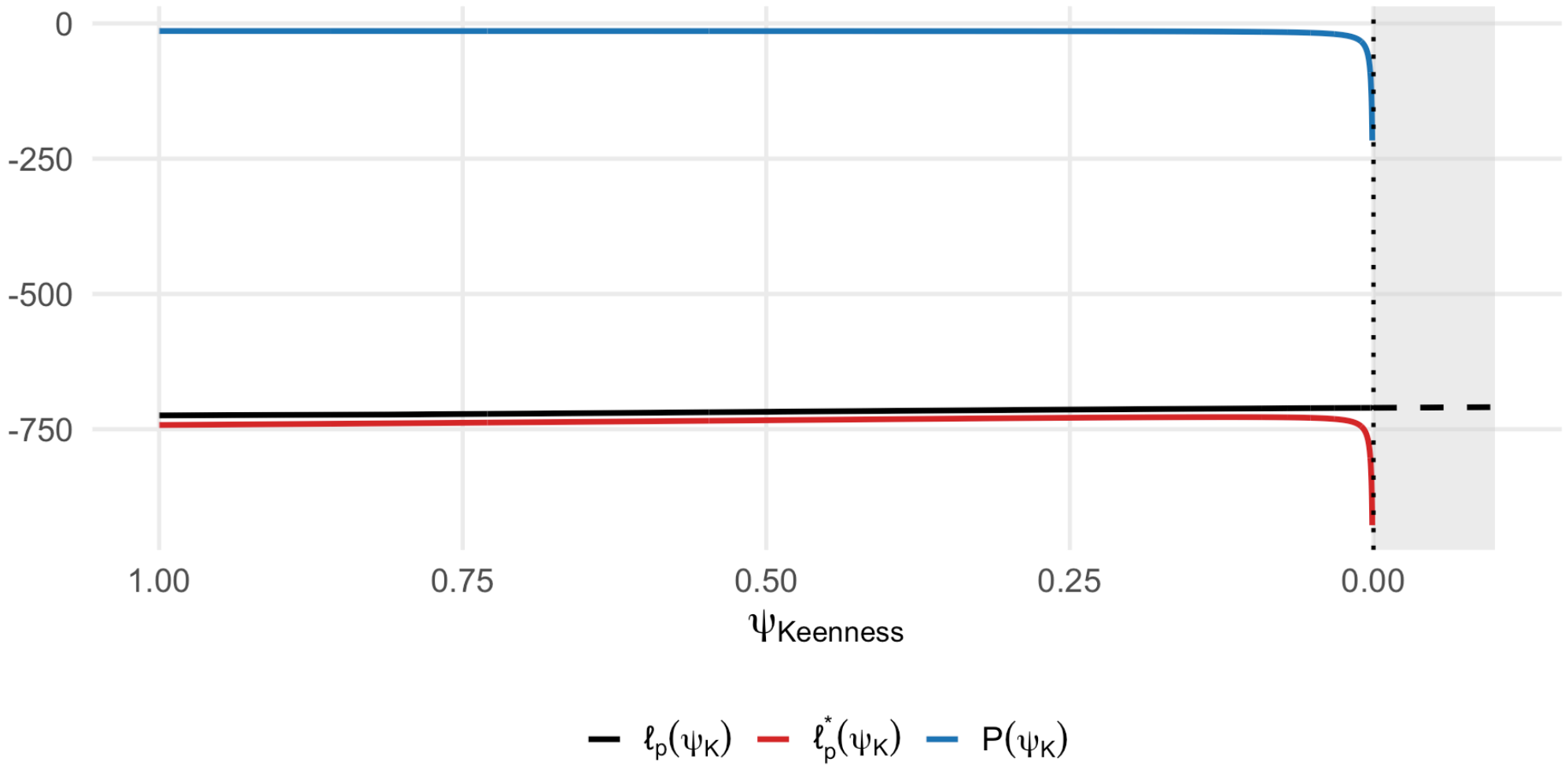
$$\ell^*(\Lambda, \Psi; S) = \ell(\Lambda\Lambda^\top + \Psi; S) + P(\Lambda, \Psi)$$

Maximum penalized likelihood (MPL) estimator

$$\bar{\Lambda}, \bar{\Psi} \in \arg\max \ell^*(\Lambda, \Psi; S) \tag{1}$$

Maximum penalized likelihood

Discouraging dominating parameter sequences



Existence

Assume:

- S is full rank
- $P(\Lambda, \Psi)$ is continuous and bounded from above
- $P(\Lambda_n, \Psi_n) \to -\infty$ for any convergent sequence (Λ_n, Ψ_n) with positive uniquenesses such that

$$\lim_{n\to\infty} \min_i [\Psi_n]_{ii} = 0, \qquad \text{when } \lim_{n\to\infty} \lambda_{\min}(\Lambda_n \Lambda_n^\top + \Psi_n) > 0$$

Then the MPL estimates $\bar{\Lambda}, \bar{\Psi}$ exist, i.e.

$$\arg\max_{\Lambda, \Psi : \Psi_{ii} > 0} \ell^*(\Lambda, \Psi; S) \neq \emptyset\,.$$

Soft penalisation

Penalisation distorts MPL estimates away from ML estimates

Scale penalty by sequence c_n to make $c_n P(\cdot, \cdot)$ asymptotically negligible

Maximum softly-penalized likelihood (MSPL) estimator

$$\bar{\Lambda}, \bar{\Psi} \in \arg\max \ell(\Lambda\Lambda^\top + \Psi; S) + c_n P(\Lambda, \Psi)$$

Consistency

Assume:

- The set of MPL estimates is nonempty
- $P(\Lambda, \Psi)$ is non-positive
- $\Sigma_0 = \Lambda_0 \Lambda_0^\top + \Psi_0$ is strongly identifiable in that for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|||\Sigma_0 - \Sigma|||_{\max} < \delta \implies |||\Lambda_0 - \Lambda Q|||_{\max} < \epsilon, \; |||\Psi_0 - \Psi|||_{\max} < \epsilon \,,$$

  for some $q \times q$ orthogonal matrix Q.

For any $\epsilon > 0$, there exists a $\delta > 0$, such that

$$\begin{matrix} |||\Sigma_0 - S|||_{\max} < \delta \\ |n^{-1} c_n P(\Lambda, \Psi)| < \delta \end{matrix} \implies \begin{matrix} |||\Lambda_0 - \bar{\Lambda} Q|||_{\max} < \epsilon \\ |||\Psi_0 - \bar{\Psi}|||_{\max} < \epsilon \end{matrix},$$

for some $q \times q$ orthogonal matrix Q.

Consistency

For any $\epsilon > 0$, there exists a $\delta > 0$, such that

$$\begin{matrix} |||\Sigma_0 - S|||_{\max} < \delta \\ |n^{-1}c_n P(\Lambda, \Psi)| < \delta \end{matrix} \quad \Longrightarrow \quad \begin{matrix} |||\Lambda_0 - \bar{\Lambda}Q|||_{\max} < \epsilon \\ |||\Psi_0 - \bar{\Psi}|||_{\max} < \epsilon \end{matrix},$$

for some $q \times q$ orthogonal matrix Q.

If $S \longrightarrow \Sigma_0$ and $P(\Lambda, \Psi)$ is an $O(1)$ function, then $c_n = o(n)$ is sufficient for consistency

Candidate penalties

Choosing c_n

Use the independence model as calibration benchmark: $\sqrt{\frac{n}{2}} \left(\frac{\hat{\Psi}_{jj}^2}{\Psi_{jj}^2} - 1 \right) \xrightarrow{d} \mathrm{N}(0, 1)$

Root-inverse-information scale is $\sqrt{2/n}$

Akaike (1987)

$$P(\Lambda, \Psi) = -\frac{\rho n}{2} \mathrm{tr} \left(\Psi^{-1/2} \Lambda \Lambda^\top \Psi^{-1/2} \right)$$

Hirose et al. (2011)

$$P(\Psi) = -\frac{\rho n}{2} \mathrm{tr} \left(\Psi^{-1/2} S \Psi^{-1/2} \right)$$

For MSPL, take $\rho = 2\sqrt{2/n^3}$.

Simulation

Setup

Based on Cooperman & Waller (2022)

- $q = 3$
- $n \in \{50, 100, 400\}$
- item-to-factor ratio $3:1, 5:1, 8:1$

Methods

Method	Figure label	Scale
ML	None	–
MPL	Akaike[n], Hirose[n]	$\rho = 1$
MSPL	Akaike[$n^{-1/2}$], Hirose[$n^{-1/2}$]	$\rho = 2\sqrt{2/n^3}$

Setup

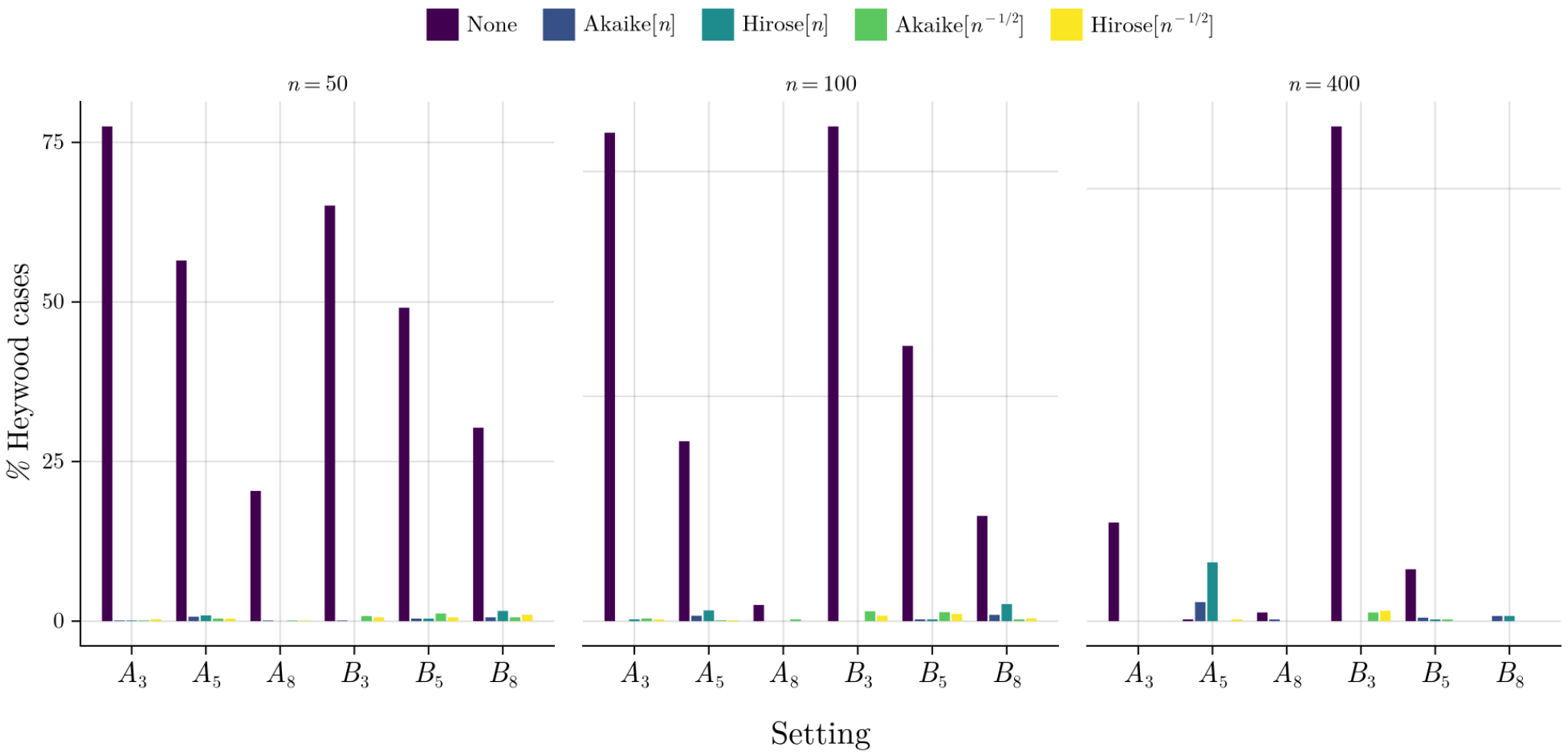
Loading Matrix A

Items	Loading 1	Loading 2	Loading 3
1	0.80	0	0
2	0.65	0	0
3	0.45	0	0
4	0	0.80	0
5	0	0.65	0
6	0	0.45	0
7	0	0	0.80
8	0	0	0.65
9	0	0	0.45

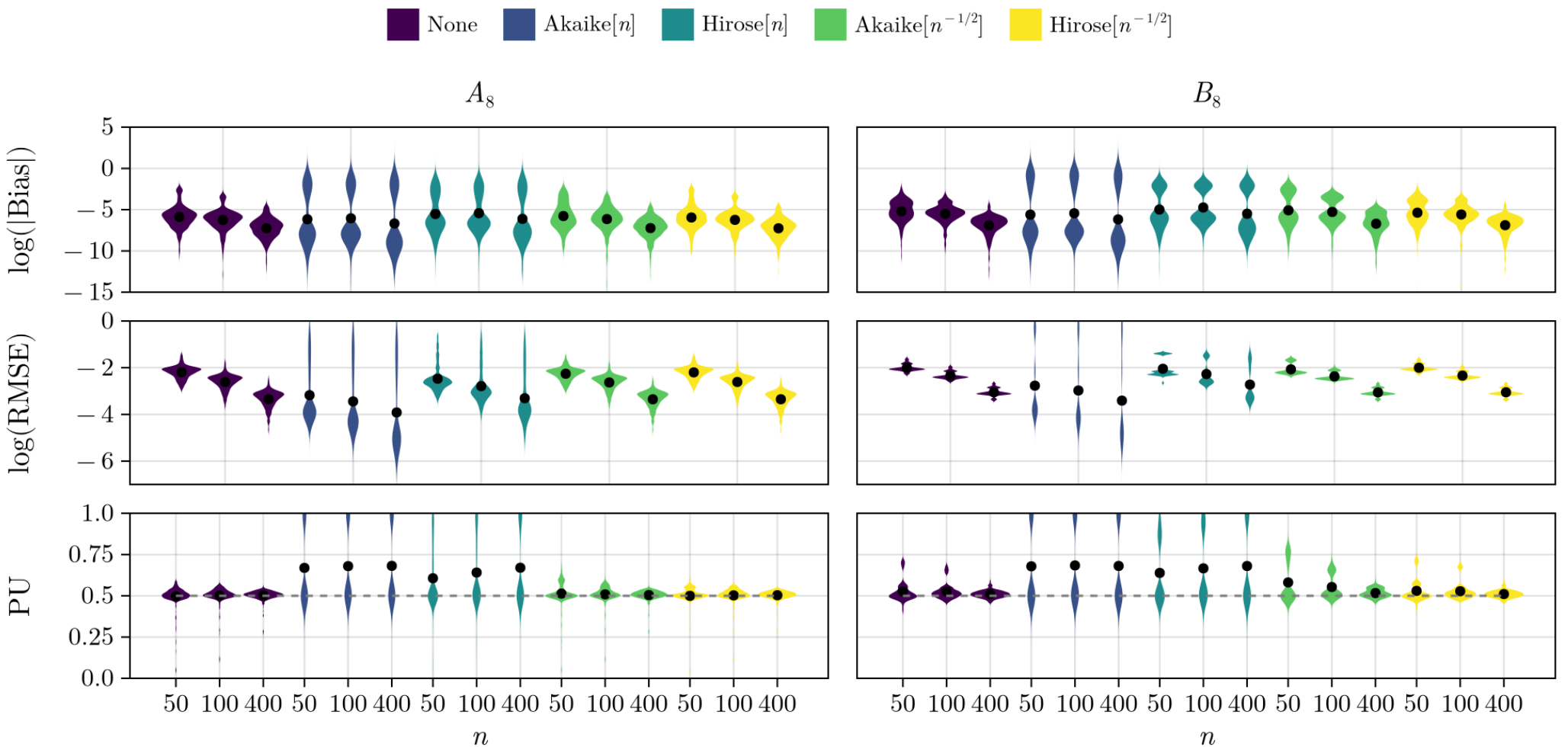
Loading Matrix B

Items	Loading 1	Loading 2	Loading 3
1	0.80	0	0
2	0.80	0	0
3	0.80	0	0
4	0	0.80	0
5	0	0.80	0
6	0	0.80	0
7	0	0	0.30
8	0	0	0.30
9	0	0	0.30

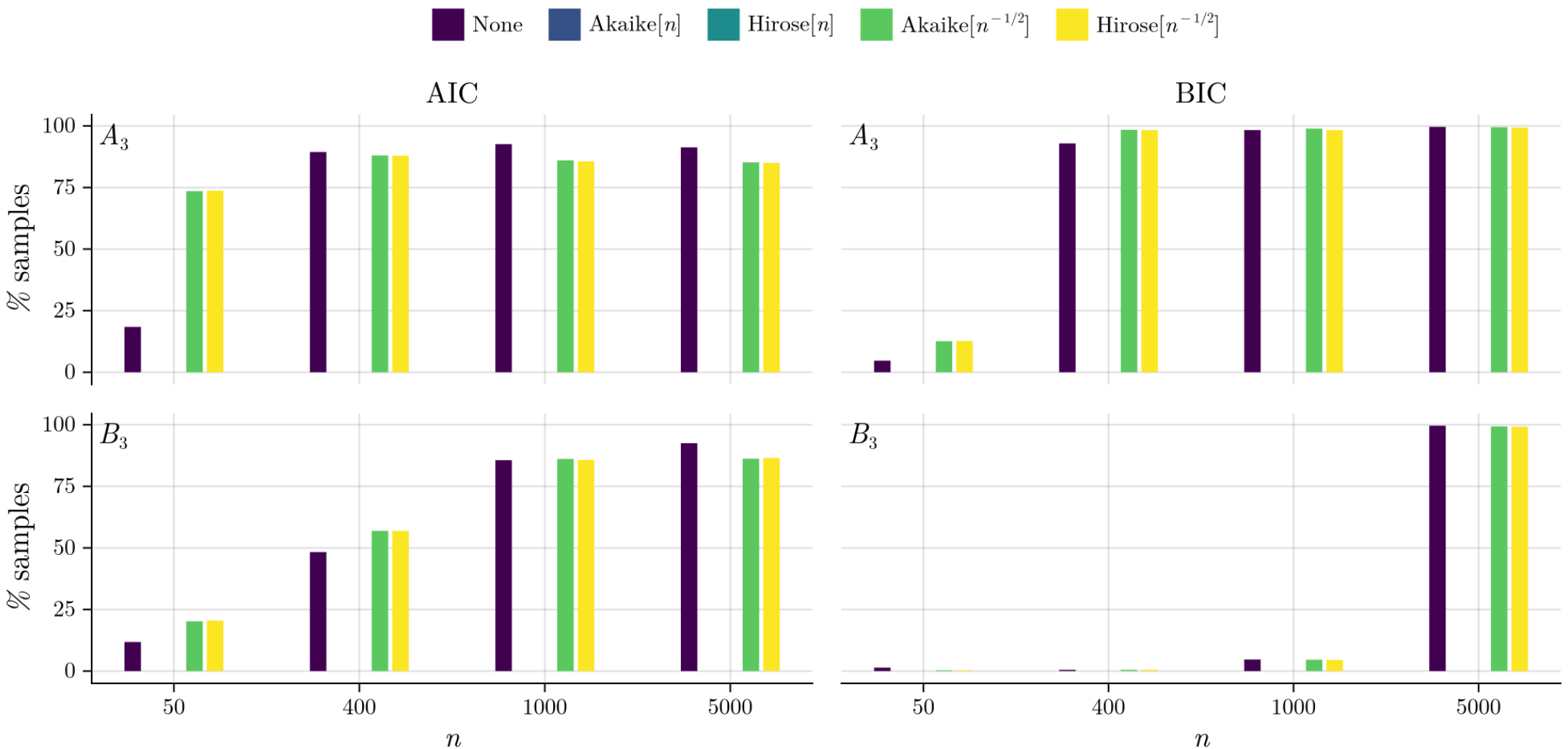
% Heywood cases



Frequentist performance



Model selection



References

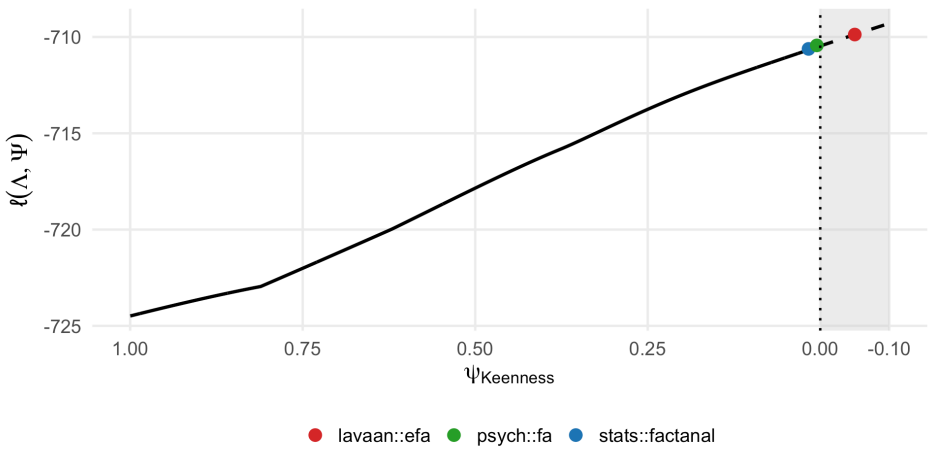
Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332. https://doi.org/10.1007/BF02294359

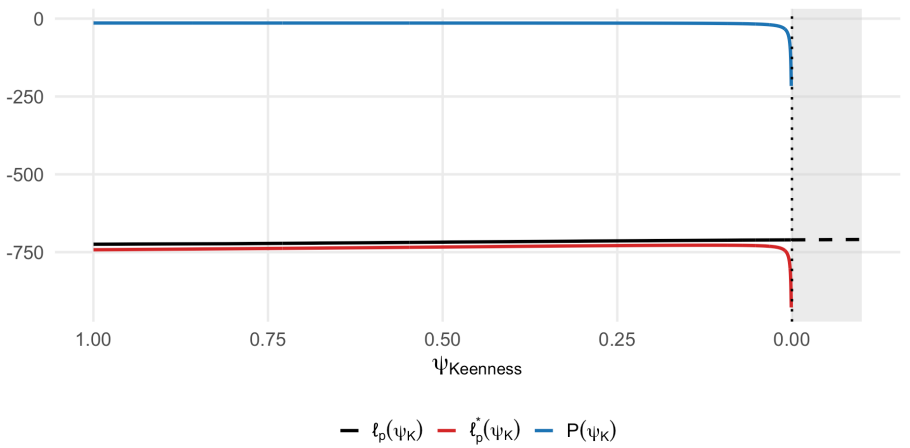
Cooperman, A. W., & Waller, N. G. (2022). Heywood you go away! Examinig causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods*, *27*(2), 156–176. https://doi.org/10.1037/met0000384

Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society, A*, *134*, 486–510. https://doi.org/10.1098/rspa.1931.0209

Hirose, K., Kawano, S., Konishi, S., & Ichikawa, M. (2011). Bayesian information criterion and selection of the number of factors in factor analysis models. *Journal of Data Science*, *9*(2), 243–259. https://doi.org/10.6339/JDS.201104_09(2).0007

Kendall, M. G. (1980). *Multivariate analysis*. Charles Griffin & Co. Ltd.





Maximum softly-penalised likelihood framework is a robust alternative to ML estimation in factor analysis

- Penalisation guarantees existence of estimates
- Soft scaling preserves ML asymptotics

Sterzinger P., Kosmidis I., and Moustaki I. (2026). *Maximum Softly Penalized Likelihood in Factor Analysis*, Psychometrika. DOI:10.1017/psy.2026.10092