# A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts

## Short user guide to the Matlab codes

Alwin Stegeman [†]

January 27, 2016

In Stegeman (2016) the following two data factor models are considered:

$$\text{Data factor model I}: \quad \min_{\substack{N^{-1}[\mathbf{F}\ \mathbf{E}]^T[\mathbf{F}\ \mathbf{E}]=\mathbf{I}_{R+J} \\ \mathbf{F}^T\mathbf{1}_N=\mathbf{0},\ \mathbf{E}^T\mathbf{1}_N=\mathbf{0} \\ \mathbf{U}\ \text{diagonal}}} \quad \text{ssq}(\mathbf{X}-\mathbf{F}\,\mathbf{P}^T-\mathbf{E}\,\mathbf{U})\,, \tag{1}$$

$$\text{Data factor model II}: \quad \min_{\substack{N^{-1}[\mathbf{F}\ \mathbf{E}]^T[\mathbf{F}\ \mathbf{E}]=\mathbf{I}_{R+J} \\ \mathbf{F}^T\mathbf{1}_N=\mathbf{0},\ \mathbf{E}^T\mathbf{1}_N=\mathbf{0} \\ \mathbf{U}\ \text{diagonal} \\ N^{-1}\mathbf{E}^T(\mathbf{X}-\mathbf{E}\mathbf{U})=\mathbf{O}}} \quad \text{ssq}(\mathbf{X}-\mathbf{F}\,\mathbf{P}^T-\mathbf{E}\,\mathbf{U})\,. \tag{2}$$

Here, $\mathbf{X} \in \mathbb{R}^{N\times J}$ is a dataset of $N$ observations on $J$ standardized continuous variables, $\mathbf{P} \in \mathbb{R}^{J\times R}$ is the loading matrix, $\mathbf{F} \in \mathbb{R}^{N\times R}$ is the matrix of factor scores, $\mathbf{E} \in \mathbb{R}^{N\times J}$ is the matrix of standardized unique parts, and diagonal $\mathbf{U} \in \mathbb{R}^{J\times J}$ contains the unique standard deviations. Let $\mathbf{S} = N^{-1}\mathbf{X}^T\mathbf{X}$ be the observed correlation matrix.

Fitting of data factor model I is done via the Matlab command

$$[\mathtt{F},\mathtt{Fd},\mathtt{P},\mathtt{E},\mathtt{Ed},\mathtt{U},\mathtt{fptot},\mathtt{mincorr}] = \mathtt{DFM1}(\mathtt{X},\mathtt{R},\mathtt{1})\,,$$

where `Fd` and `Ed` denotes the determinate parts of $\mathbf{F}$ and $\mathbf{E}$, respectively, `fptot` is the fit percentage of the model, and `mincorr` contains the minimal correlations for each factor. The third input argument in the function `DFM1` indicates whether output on various fit measures should be given or not.

Analogously, data factor model II is fitted via the Matlab command

$$[\mathtt{F},\mathtt{Fd},\mathtt{P},\mathtt{E},\mathtt{Ed},\mathtt{U},\mathtt{comm},\mathtt{ecv},\mathtt{ecv\_total},\mathtt{fptot},\mathtt{mincorr}] = \mathtt{DFM2}(\mathtt{X},\mathtt{R},\mathtt{1})\,,$$

[†] The author is with the Heijmans Institute for Psychological Research, University of Groningen, The Netherlands, email: a.w.stegeman@rug.nl, URL: http://www.alwinstegeman.nl

where `comm` contains the communalities (diagonal of $\mathbf{S} - \mathbf{U}^2$), `ecv` contains the percentages of explained common variance for each variable, and `ecv_total` is the overall percentage of explained common variance. The algorithm for data factor model II uses MRFA to estimate $\mathbf{U}$. This is done by the Matlab function `mrfa.m`.

Also included is the file `data_Grice.m` that contains the dataset of Grice (2001) that is analyzed in Stegeman (2016). The file also contains the commands to fit data factor model II to this dataset, perform a Varimax rotation, and to obtain factor score densities due to indeterminacy under model II. The latter are plotted by the function `factor_score_density.m`.

# References

Grice, J.W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*, 430–450.

Stegeman, A. (2016). A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts. *Computational Statistics and Data Analysis*, to appear. Preprint available online at `www.alwinstegeman.nl`