

1 Supplementary Methods

2 *Gradient function of ACMTF*

3 In the case where the first mode is shared across all data blocks, and assuming that every term in
 4 $f(\alpha, \beta, \varepsilon, \Lambda, \mathbf{A}, \mathbf{B}^{(1)}, \mathbf{C}^{(1)}, \dots, \mathbf{B}^{(P)}, \mathbf{C}^{(P)})$ is multiplied by 1/2, the gradient of the loss function ∇f
 5 with size $F(I + J^{(1)} + K^{(1)} + \dots + J^{(P)} + K^{(P)} + P) \times 1$ becomes

$$6 \quad \nabla f = \left[\text{vec} \left(\frac{\partial f}{\partial \mathbf{A}} \right)^T \text{vec} \left(\frac{\partial f}{\partial \mathbf{B}^{(1)}} \right)^T \text{vec} \left(\frac{\partial f}{\partial \mathbf{C}^{(1)}} \right)^T \dots \text{vec} \left(\frac{\partial f}{\partial \mathbf{B}^{(P)}} \right)^T \text{vec} \left(\frac{\partial f}{\partial \mathbf{C}^{(P)}} \right)^T \frac{\partial f}{\partial \lambda_1} \frac{\partial f}{\partial \lambda_2} \dots \frac{\partial f}{\partial \lambda_P} \right]^T \quad (5)$$

7 where

$$8 \quad \frac{\partial f}{\partial \mathbf{A}} = \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\lambda_p^T \odot \mathbf{C}^{(p)} \odot \mathbf{B}^{(p)}) + \alpha (\mathbf{A} - \bar{\mathbf{A}}),$$

$$9 \quad \frac{\partial f}{\partial \mathbf{B}^{(p)}} = \sum_{p=1}^P (\hat{\mathbf{X}}_b^{(p)} - \mathbf{X}_b^{(p)}) (\lambda_p^T \odot \mathbf{C}^{(p)} \odot \mathbf{A}) + \alpha (\mathbf{B}^{(p)} - \bar{\mathbf{B}}^{(p)}),$$

$$10 \quad \frac{\partial f}{\partial \mathbf{C}^{(p)}} = \sum_{p=1}^P (\hat{\mathbf{X}}_c^{(p)} - \mathbf{X}_c^{(p)}) (\lambda_p^T \odot \mathbf{B}^{(p)} \odot \mathbf{A}) + \alpha (\mathbf{C}^{(p)} - \bar{\mathbf{C}}^{(p)}),$$

$$11 \quad \frac{\partial f}{\partial \lambda_{pf}} = (\hat{\mathbf{X}}^{(p)} - \mathbf{X}^{(p)}) \times_1 \mathbf{a}_f \times_2 \mathbf{b}_f^{(p)} \times_3 \mathbf{c}_f^{(p)} + \frac{\beta}{2} \frac{\lambda_{pf}}{\sqrt{\lambda_{pf}^2 + \varepsilon}}$$

12 and \times_n is the tensor-vector product in the n -th mode as defined in [35]. $\bar{\mathbf{A}}$ corresponds to the matrix
 13 \mathbf{A} with columns divided by their 2-norms. The derivation of this gradient is reported in [18]. The
 14 `acmtf_opt()` function runs this model in the provided R package.

15 *Gradient function of ACMTF-R*

16 Assuming that the first mode is shared across all data blocks and that every term in
 17 $f(\alpha, \beta, \varepsilon, \pi, \Lambda, \mathbf{A}, \mathbf{B}^{(1)}, \mathbf{C}^{(1)}, \dots, \mathbf{B}^{(P)}, \mathbf{C}^{(P)})$ is multiplied by 1/2, the gradient of the loss function
 18 becomes

$$19 \quad \nabla f = \left[\text{vec} \left(\frac{\partial f}{\partial \mathbf{A}} \right)^T \text{vec} \left(\frac{\partial f}{\partial \mathbf{B}^{(1)}} \right)^T \text{vec} \left(\frac{\partial f}{\partial \mathbf{C}^{(1)}} \right)^T \dots \text{vec} \left(\frac{\partial f}{\partial \mathbf{B}^{(P)}} \right)^T \text{vec} \left(\frac{\partial f}{\partial \mathbf{C}^{(P)}} \right)^T \frac{\partial f}{\partial \lambda_1} \frac{\partial f}{\partial \lambda_2} \dots \frac{\partial f}{\partial \lambda_P} \right]^T \quad (8)$$

20 where

$$21 \quad \frac{\partial f}{\partial \mathbf{A}} = \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p^T \odot \mathbf{C}^{(p)} \odot \mathbf{B}^{(p)}) + (1 - \pi) (\mathbf{A} \boldsymbol{\rho} - \mathbf{y}) \boldsymbol{\rho}^T + \alpha (\mathbf{A} - \bar{\mathbf{A}}),$$

$$22 \quad \frac{\partial f}{\partial \mathbf{B}^{(p)}} = \pi \sum_{p=1}^P (\hat{\mathbf{X}}_b^{(p)} - \mathbf{X}_b^{(p)}) (\boldsymbol{\lambda}_p^T \odot \mathbf{C}^{(p)} \odot \mathbf{A}) + \alpha (\mathbf{B}^{(p)} - \bar{\mathbf{B}}^{(p)}),$$

$$23 \quad \frac{\partial f}{\partial \mathbf{C}^{(p)}} = \pi \sum_{p=1}^P (\hat{\mathbf{X}}_c^{(p)} - \mathbf{X}_c^{(p)}) (\boldsymbol{\lambda}_p^T \odot \mathbf{B}^{(p)} \odot \mathbf{A}) + \alpha (\mathbf{C}^{(p)} - \bar{\mathbf{C}}^{(p)}),$$

$$24 \quad \frac{\partial f}{\partial \lambda_{pf}} = (\hat{\mathbf{X}}^{(p)} - \mathbf{X}^{(p)}) \times_1 \mathbf{a}_f \times_2 \mathbf{b}_f^{(p)} \times_3 \mathbf{c}_f^{(p)} + \frac{\beta}{2} \frac{\lambda_{pf}}{\sqrt{\lambda_{pf}^2 + \epsilon}}$$

25 and \times_n is the tensor-vector product in the n -th mode as defined in [35]. $\bar{\mathbf{A}}$ corresponds to the matrix
 26 \mathbf{A} with the columns divided by their 2-norms. The derivation of this gradient is reported in
 27 **Appendix 1**. The `acmtfr_opt()` function runs this algorithm in the provided R package.

28 *Missing data*

29 Missing data is an unfortunate reality of sampling biological data, especially for microbiome data.
 30 In cohort studies, some participants might not turn up for sampling at one or more time points, or
 31 they may drop out altogether. Low library sizes may cause some samples to be removed from the
 32 data. Such missing data ends up as rows of missing data in the three-way array. Fortunately, the
 33 ACMTF and ACMTF-R function can be adapted to accommodate this. This section highlights the
 34 mathematical adaptations required for ACMTF and ACMTF-R to work appropriately in the
 35 presence of missing data. First, the definitions of the binary weight array \mathbf{W} and the weighted norm
 36 $\|\cdot\|_{\mathbf{W}}$ are given. Then the modifications for ACMTF and ACMTF-R are outlined.

37 *Binary weight array \mathbf{W}*

38 All of the mathematical adaptations presented in this section required a binary three-way array of
 39 weights \mathbf{W} encoding the non-missing values in the data \mathbf{X} where

$$40 \quad w_{ijk} = \begin{cases} 0 & \text{if } x_{ijk} \text{ is missing} \\ 1 & \text{otherwise,} \end{cases} \quad (1)$$

41 and the array element x_{ijk} in \mathbf{X} corresponding to subject i and feature j at time point k is used to
 42 determine the corresponding array element w_{ijk} in \mathbf{W} and a two-way array of weights is defined
 43 equivalently [1, 2].

44 *Weighted norm*45 The weighted norm of a three-way array $\underline{\mathbf{X}}$ is defined as

46
$$\|\underline{\mathbf{X}}\|_W = \|\underline{\mathbf{W}} * \underline{\mathbf{X}}\| \quad (2)$$

47 where the weight array $\underline{\mathbf{W}}$ is used to encode the non-missing array elements in $\underline{\mathbf{X}}$ [1, 2]. An
48 equivalent definition is used for the two-way and vector cases.49 *Weighted ACMTF*

50 In ACMTF, the weighted calculation of the loss only affects the first term

$$\begin{aligned}
f(\alpha, \beta, \epsilon, \Lambda, \mathbf{A}, \mathbf{B}^{(1)}, \mathbf{C}^{(1)}, \dots, \mathbf{B}^{(P)}, \mathbf{C}^{(P)}) &= \sum_{p=1}^P \|\underline{\mathbf{X}}^{(p)} - \hat{\underline{\mathbf{X}}}^{(p)}\|_W^2 \\
&+ \beta \sum_{p=1}^P \sum_{f=1}^F \sqrt{\lambda_{fp}^2 + \epsilon} \\
&+ \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{a}_f\| - 1)^2 \\
&+ \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{b}_f^{(p)}\| - 1)^2 \\
&+ \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{c}_f^{(p)}\| - 1)^2
\end{aligned} \quad (3)$$

52 and the gradient function is unaffected [3].

53 *Weighted ACMTF-R*

54 In ACMTF-R, the weighted calculation of the loss only affects the first term

$$\begin{aligned}
f(\alpha, \beta, \varepsilon, \pi, \mathbf{A}, \mathbf{B}^{(1)}, \mathbf{C}^{(1)}, \dots, \mathbf{B}^{(P)}, \mathbf{C}^{(P)}) = & \pi \sum_{p=1}^P \|\underline{\mathbf{X}}^{(p)} - \widehat{\underline{\mathbf{X}}}^{(p)}\|_W^2 \\
& + (1 - \pi) \|\mathbf{y} - \mathbf{A}\boldsymbol{\rho}\|^2 \\
& + \beta \sum_{p=1}^P \sum_{f=1}^F \sqrt{\lambda_{fp}^2 + \varepsilon} \\
& + \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{a}_f\| - 1)^2 \\
& + \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{b}_f^{(p)}\| - 1)^2 \\
& + \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{c}_f^{(p)}\| - 1)^2
\end{aligned} \tag{4}$$

55

56 and the gradient function is unaffected (**Appendix 1**).

57 When y_{new} is calculated for a new sample \mathbf{x}_{new} containing some missing values, the weighted
58 calculation of the joint scores becomes

59

$$\mathbf{a}_{new} = \mathbf{W} * \mathbf{Z}^+ \text{vec}(x_{new})$$

60 where \mathbf{Z}^+ is the Moore-Penrose inverse of \mathbf{Z} and \mathbf{W} is a binary weight matrix encoding the non-
61 missing data for $\text{vec}(x_{new})$. Subsequently y_{new} is found by using the regression coefficients $\boldsymbol{\rho}$ as in
62 the non-weighted version of the procedure.

63 *Model evaluation - Fit*

64 Model evaluation is typically done using the percentage of variance explained. In the context of
65 ACMTF-R this can be done for a data block $\underline{\mathbf{X}}^{(p)}$ and \mathbf{y} through

66

$$Fit_{\underline{\mathbf{X}}^{(p)}}(\%) = \left(1 - \frac{\|\underline{\mathbf{X}}^{(p)} - \widehat{\underline{\mathbf{X}}}^{(p)}\|^2}{\|\underline{\mathbf{X}}^{(p)}\|^2} \right) * 100 \tag{10}$$

67 and

68

$$Fit_{\mathbf{y}}(\%) = \left(1 - \frac{\|\mathbf{y} - \widehat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} \right) * 100. \tag{11}$$

69 While variance explained gives an overall impression of the model fit, it does not give a good
70 indication of whether the input loadings have been found, nor if the correct CLD structure has been
71 found, especially in the context of noise.

72 When missing data is present, the fit can be computed for a data block $\underline{\mathbf{X}}^{(p)}$ through

73
$$Fit_{\underline{\mathbf{X}}^{(p)}}(\%) = \left(1 - \frac{\|\mathbf{X}^{(p)} - \hat{\mathbf{X}}^{(p)}\|_W^2}{\|\mathbf{X}^{(p)}\|_W^2} \right) * 100 \quad (10)$$

74 and the calculation of the variation explained in \mathbf{y} remains unchanged.

75 *Model evaluation – Lambda Similarity Index*

76 We define the Lambda Similarity Index (*LSI*) to compare the found lambda matrix $\hat{\mathbf{\Lambda}}$ to the true
77 lambda matrix $\mathbf{\Lambda}$ using an inverted sum-of-squared residuals approach

78
$$LSI = \frac{1}{1 + \sum_{p=1}^P \sum_{f=1}^F (\lambda_{pf} - \hat{\lambda}_{pf})^2} \quad (13)$$

79 where p is the block index and f is the component number. This metric has the range $(0,1]$, where
80 $LSI = 1$ corresponds to the ACMTF-R model finding the correct CLD-structure and $LSI \rightarrow 0$
81 corresponds to the ACMTF-R model not finding the correct CLD-structure in any way. LSI does
82 not take the recovered factor loadings into account.

83 *Model evaluation – Matching components*

84 Both FMS and LSI require a best matching of components due to the permutational freedom of the
85 model. However, the chosen permutation should be the same for both metrics. This is done by first
86 computing FMS and LSI for all pairwise combinations of factors and subsequently using the
87 Hungarian algorithm [4, 5] to find the matching of components that maximises both metrics. The
88 combination of FMS and LSI will be used to evaluate CLD-structure and factor recovery of ACMTF-
89 R for all simulations.

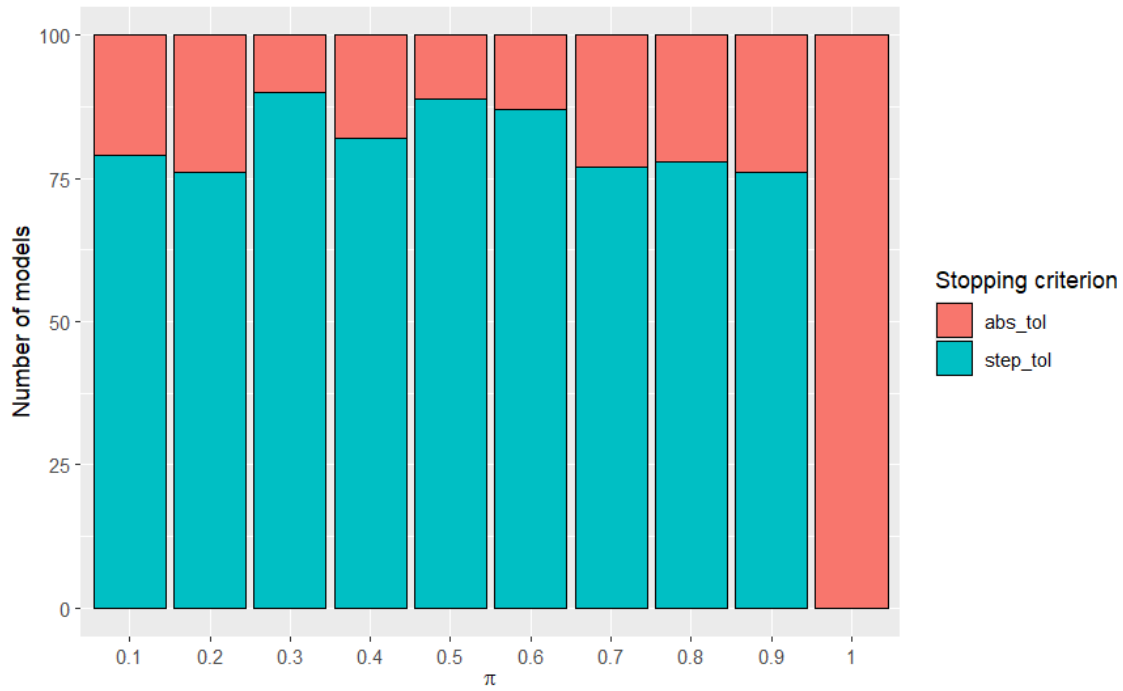
90 **Supplementary References**

- 91 1. Acar E, Kolda TG, Dunlavy DM. All-at-once Optimization for Coupled Matrix and Tensor
92 Factorizations. 2011.
- 93 2. Acar E, Dunlavy DM, Kolda TG, Mørup M. Scalable tensor factorizations for incomplete data.
94 Chemom Intell Lab Syst. 2011;106:41–56.
- 95 3. Acar E, Papalexakis EE, Gürdeniz G, Rasmussen MA, Lawaetz AJ, Nilsson M, et al. Structure-
96 revealing data fusion. BMC Bioinformatics. 2014;15:239.
- 97 4. Kuhn HW. The Hungarian method for the assignment problem. Nav Res Logist Q. 1955;2:83–
98 97.
- 99 5. Kuhn HW. Variants of the hungarian method for assignment problems. Nav Res Logist Q.
100 1956;3:253–8.

101 6. Acar E, Papalexakis EE, Gürdeniz G, Rasmussen MA, Lawaetz AJ, Nilsson M, et al. Structure-
102 revealing data fusion. BMC Bioinformatics. 2014;15:239.

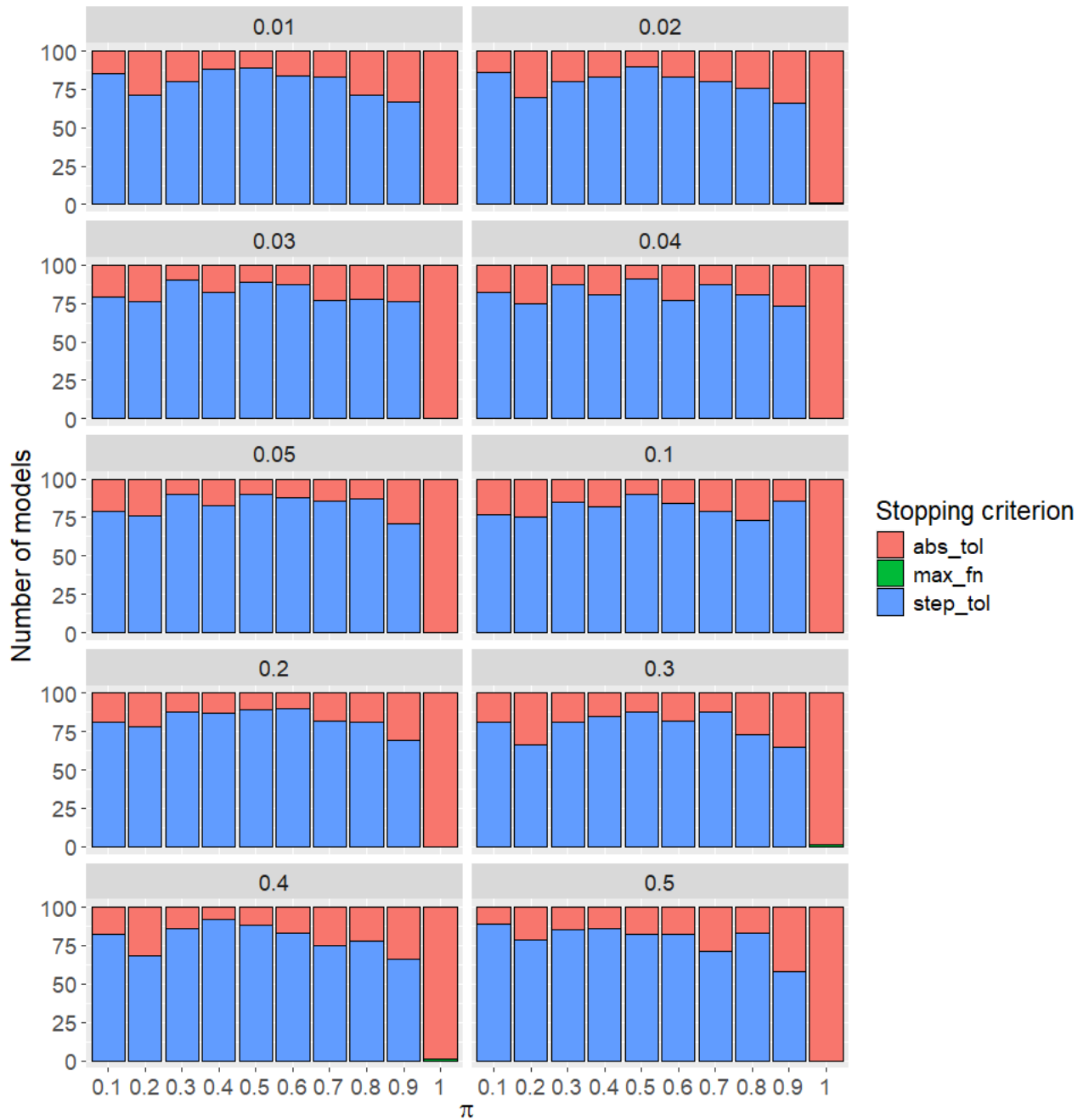
103

104

105 **Supplementary Figures**

106

107 **Supplementary Figure 1.** Overview of the stopping criterion that terminated the ACMTF-R line
108 search algorithm in the first simulation. abs_tol: minimum value of the loss function. step_tol:
109 absolute tolerance for the size of the parameter update.



110

111 **Supplementary Figure 2.** Overview of the stopping criterion that terminated the ACMTF-R line
 112 search algorithm in the second simulation. Plots are faceted by the size of the hidden component
 113 (δ). abs_tol: minimum value of the loss function. step_tol: absolute tolerance for the size of the
 114 parameter update. max_fn: maximum number of function evaluations.

115



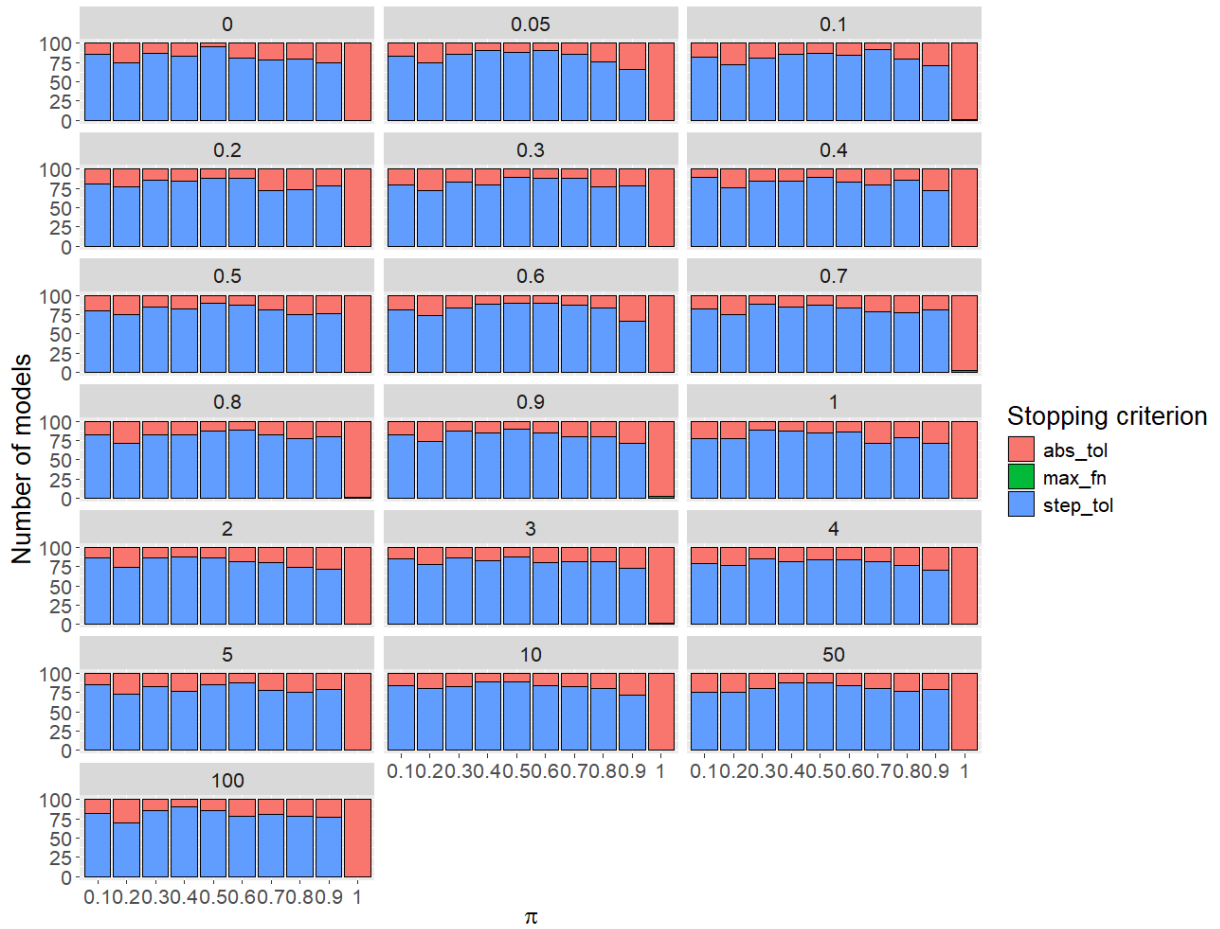
116

117 **Supplementary Figure 3.** Overview of the stopping criterion that terminated the ACMTF-R line118 search algorithm in the third simulation. Plots are faceted by the noise level on X (η_X). abs_tol:

119 minimum value of the loss function. rel_tol: relative change of the loss function. step_tol: absolute

120 tolerance for the size of the parameter update. max_fn: maximum number of function evaluations.

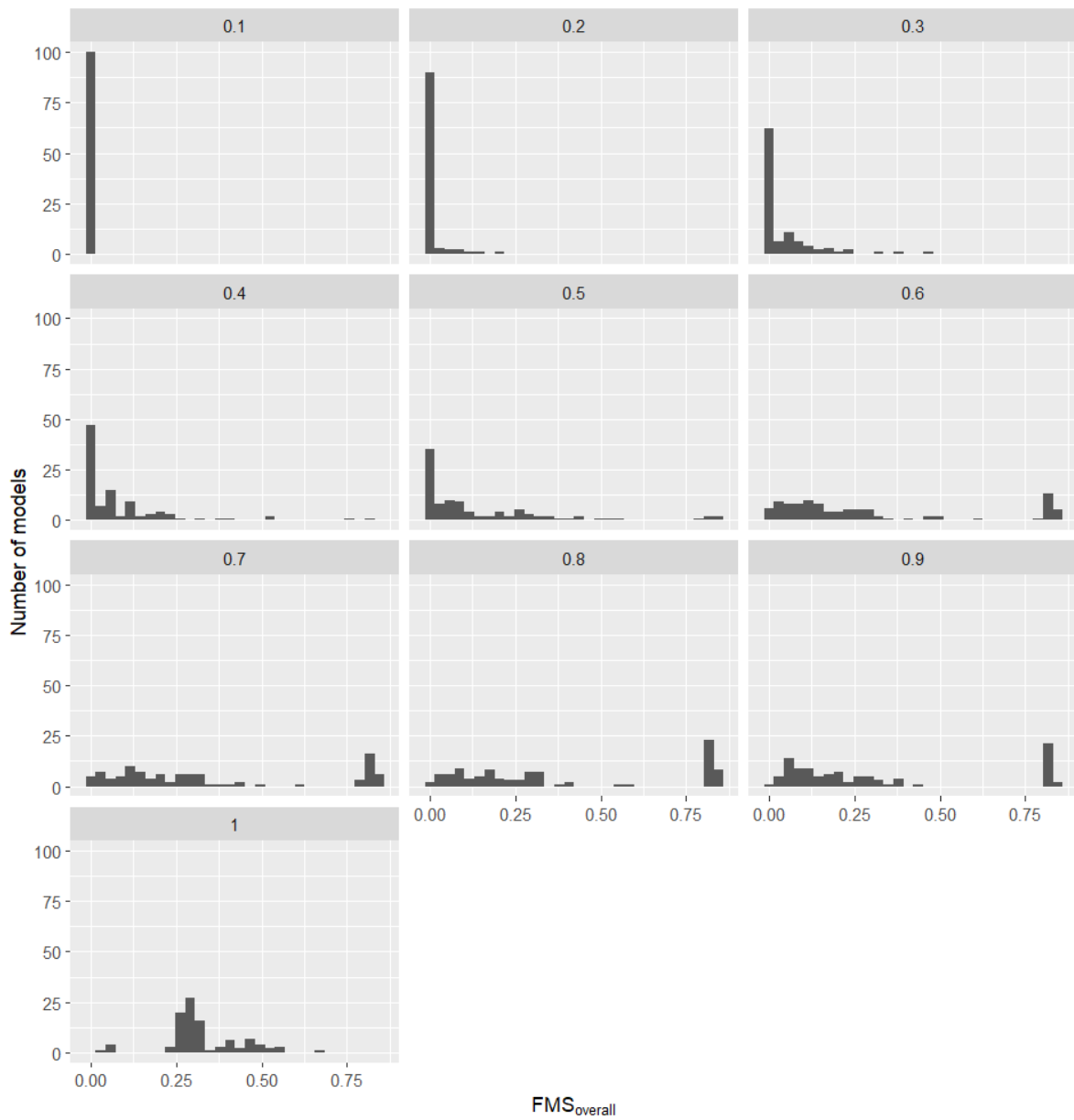
121



122

123 **Supplementary Figure 4.** Overview of the stopping criterion that terminated the ACMTF-R line
 124 search algorithm in the fourth simulation. Plots are faceted by the noise level on \mathbf{y} (η_y). abs_tol:
 125 minimum value of the loss function. step_tol: absolute tolerance for the size of the parameter update.
 126 max_fn: maximum number of function evaluations.

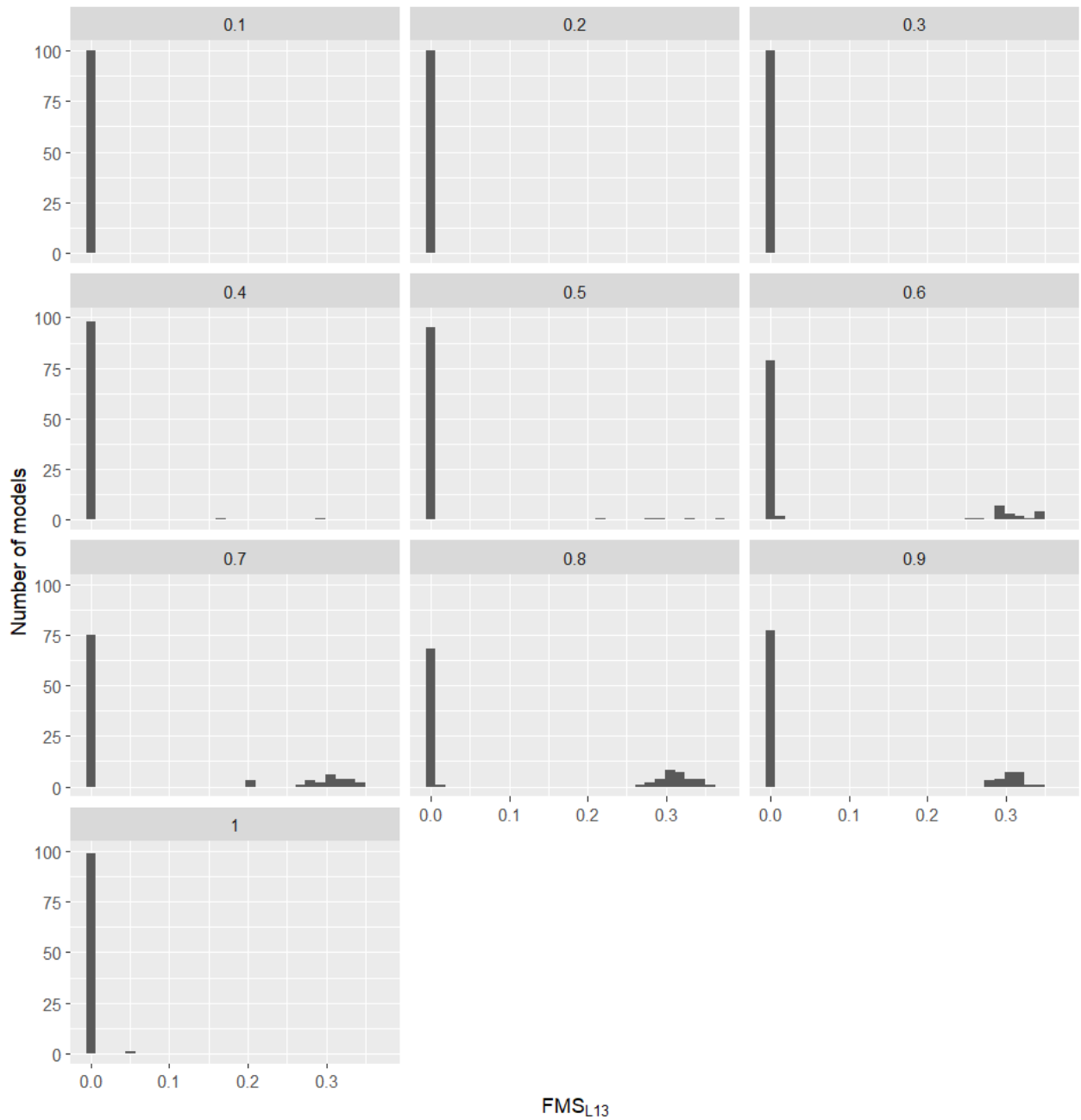
127



128

129 **Supplementary Figure 5.** Histogram of the $FMS_{overall}$ observed across the randomly initialised
 130 ACMTF and ACMTF-R models in the first simulation. Plots are faceted by π value and each contain
 131 100 randomly initialised models. ACMTF corresponds to $\pi = 1$.

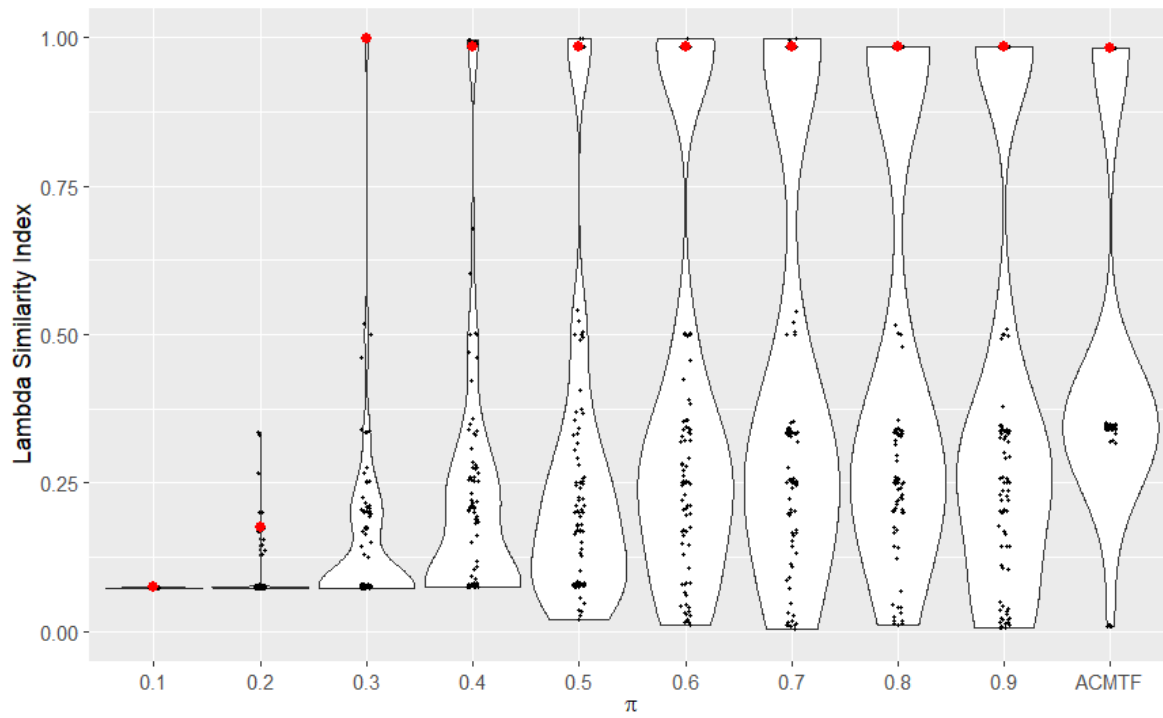
132



133

134 **Supplementary Figure 6.** Histogram of the FMS_{L13} observed across the randomly initialised ACMTF
 135 and ACMTF-R models in the first simulation. Plots are faceted by π value and each contain 100
 136 randomly initialised models. ACMTF corresponds to $\pi = 1$.

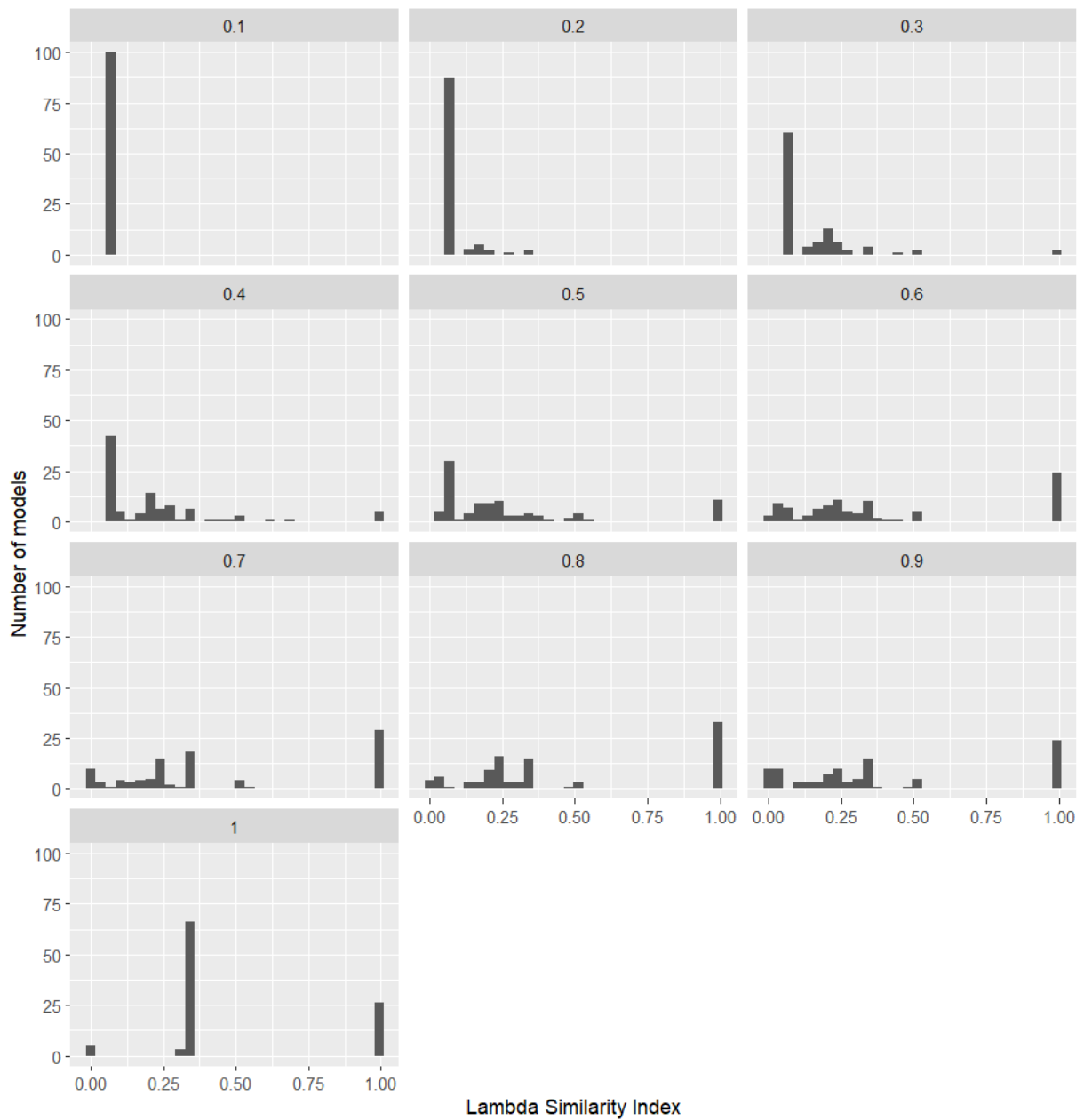
137



138

139 **Supplementary Figure 7.** Violin plots showing the Lambda Similarity Index observed across the
140 randomly initialised ACMTF and ACMTF-R models in the first simulation. The model with the
141 lowest loss per setting of π is indicated with a red data point.

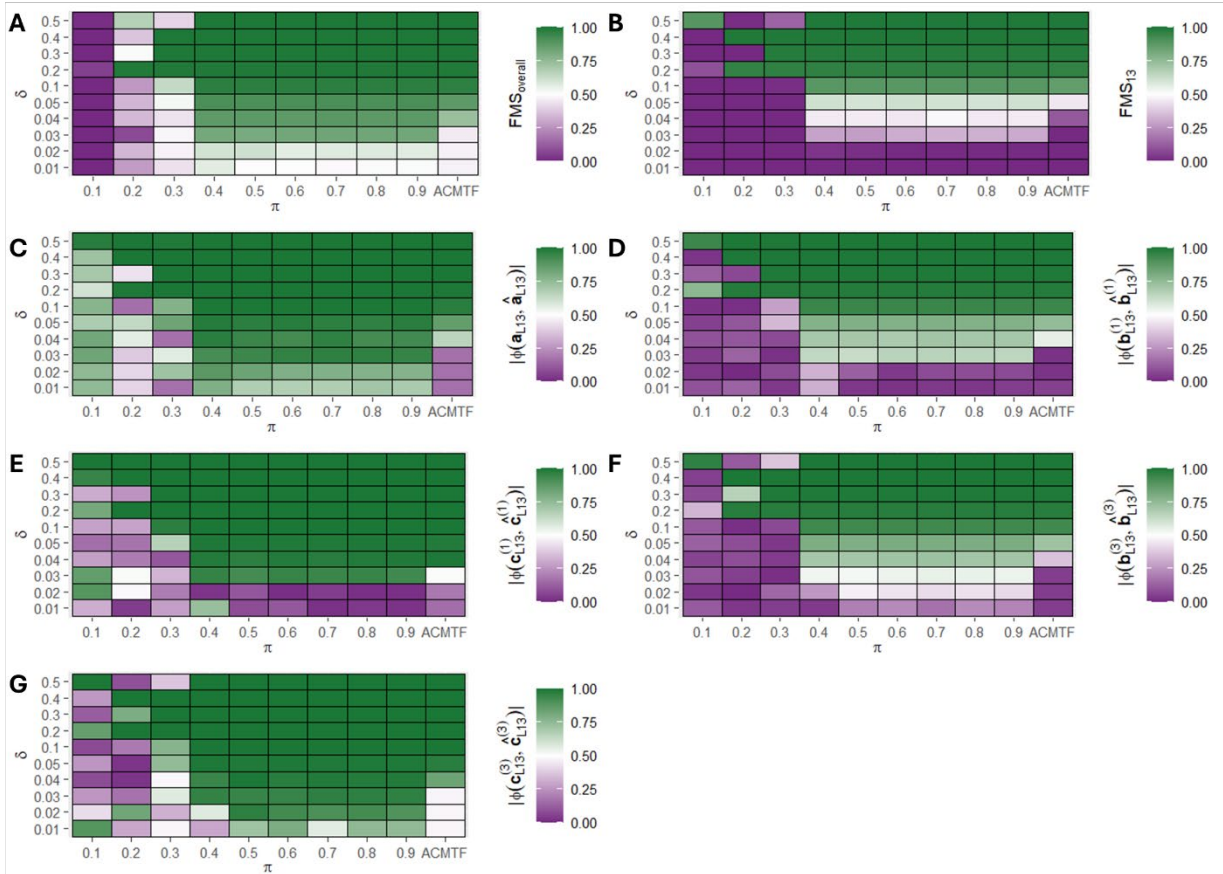
142



143

144 **Supplementary Figure 8.** Histogram of the *LSI* observed across the randomly initialised ACMTF and
 145 ACMTF-R models in the first simulation. Plots are faceted by π value and each contain 100
 146 randomly initialised models. ACMTF corresponds to $\pi = 1$.

147



148

149 **Supplementary Figure 9. Factor recovery for varying sizes of the hidden component L_{13} in ACMTF**

150 **and ACMTF-R.** One hundred randomly initialised eight-component ACMTF-R and ACMTF models were

151 fitted for each combination of hidden component size (δ) and tuning (π), after which the lowest-loss

152 models were compared. Outcome per setting for (A) overall factor recovery, (B) factor recovery of the

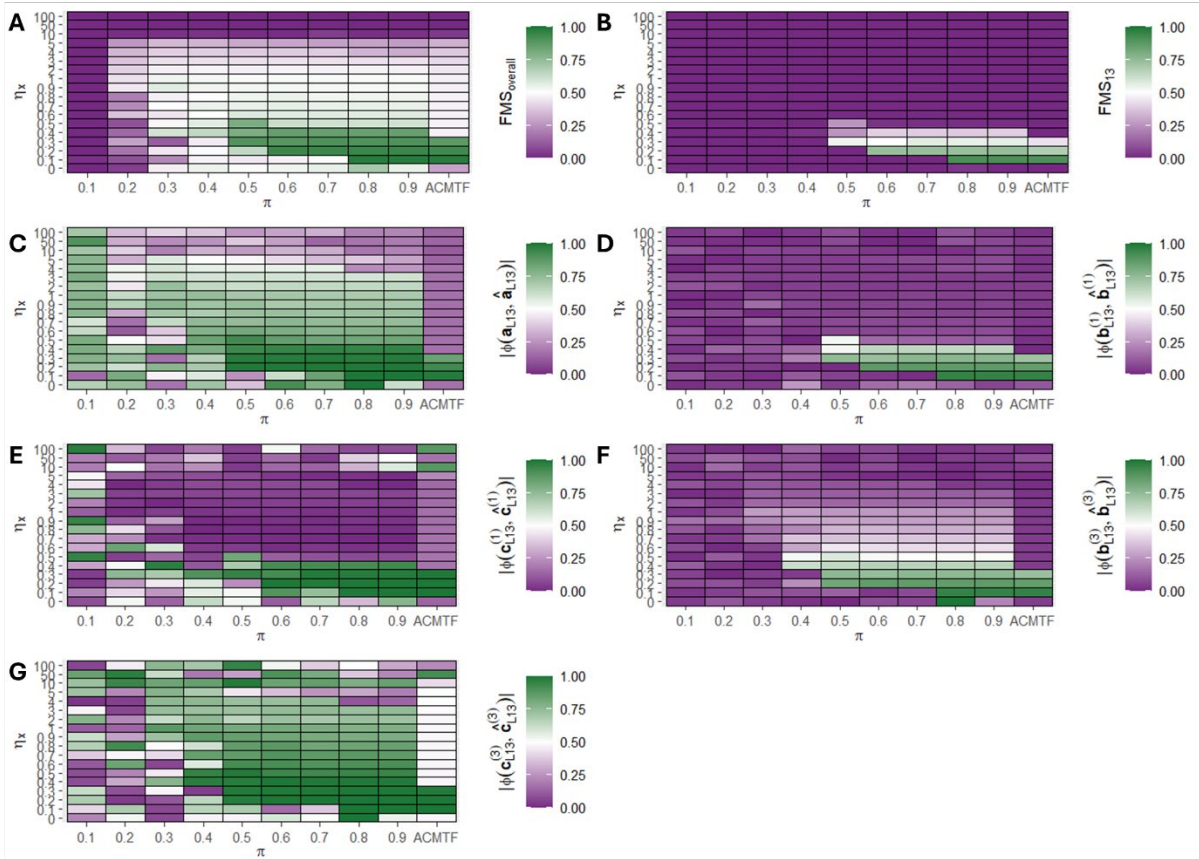
153 hidden component, (C) shared subject mode recovery of the hidden component, (D) feature mode

154 recovery of $\underline{X}^{(1)}$ for the hidden component, (E) time mode recovery of $\underline{X}^{(1)}$ for the hidden component,

155 (F) feature mode recovery of $\underline{X}^{(3)}$ for the hidden component, (G) time mode recovery of $\underline{X}^{(3)}$ for the

156 hidden component.

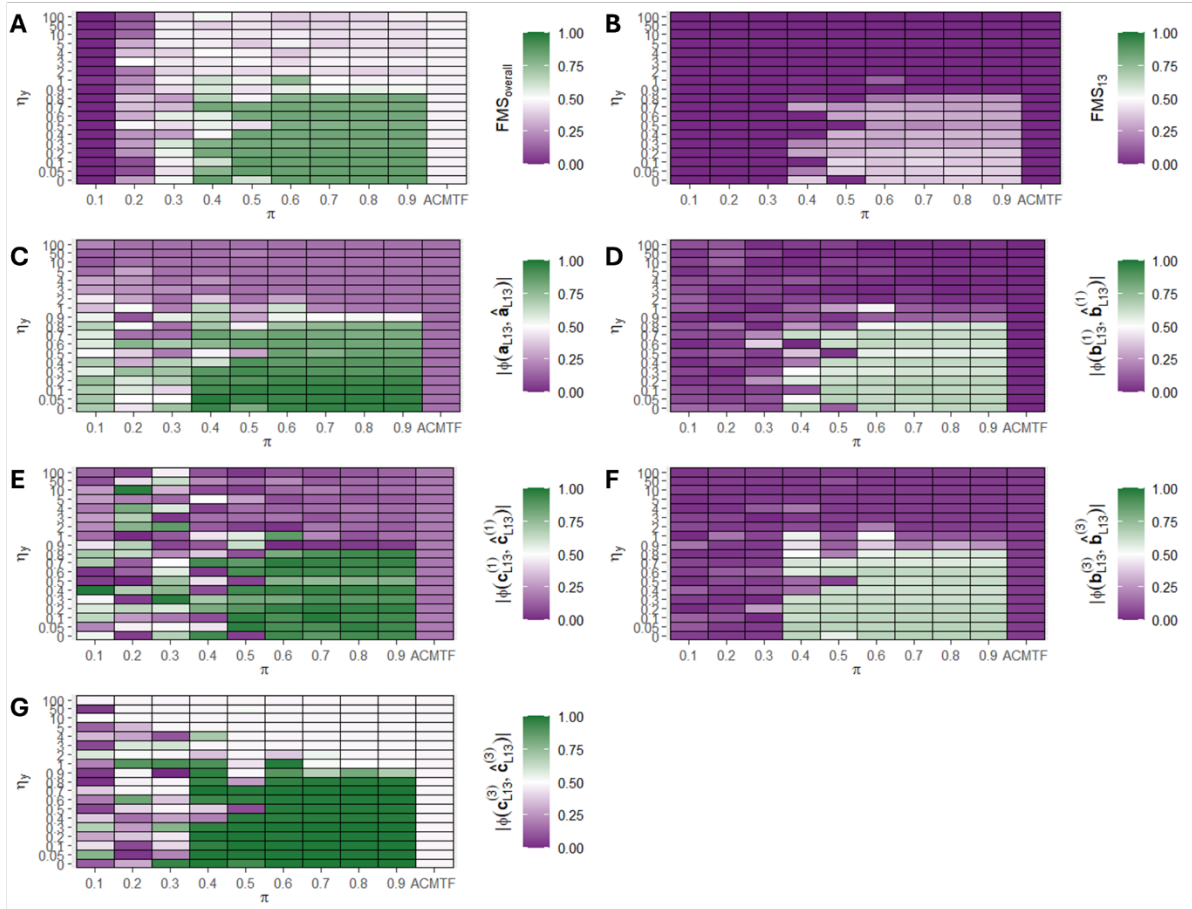
157



158

159 **Supplementary Figure 10. Factor recovery in ACMTF and ACMTF-R for varying noise levels on X.** One
 160 hundred randomly initialised eight-component ACMTF-R and ACMTF models were fitted for each
 161 combination of noise on X (η_X) and tuning parameter (π), after which the lowest-loss models were
 162 compared. Outcome per setting for (A) overall factor recovery, (B) factor recovery of the hidden
 163 component, (C) shared subject mode recovery of the hidden component, (D) feature mode recovery
 164 of $\underline{X}^{(1)}$ for the hidden component, (E) time mode recovery of $\underline{X}^{(1)}$ for the hidden component, (F)
 165 feature mode recovery of $\underline{X}^{(3)}$ for the hidden component, (G) time mode recovery of $\underline{X}^{(3)}$ for the
 166 hidden component.

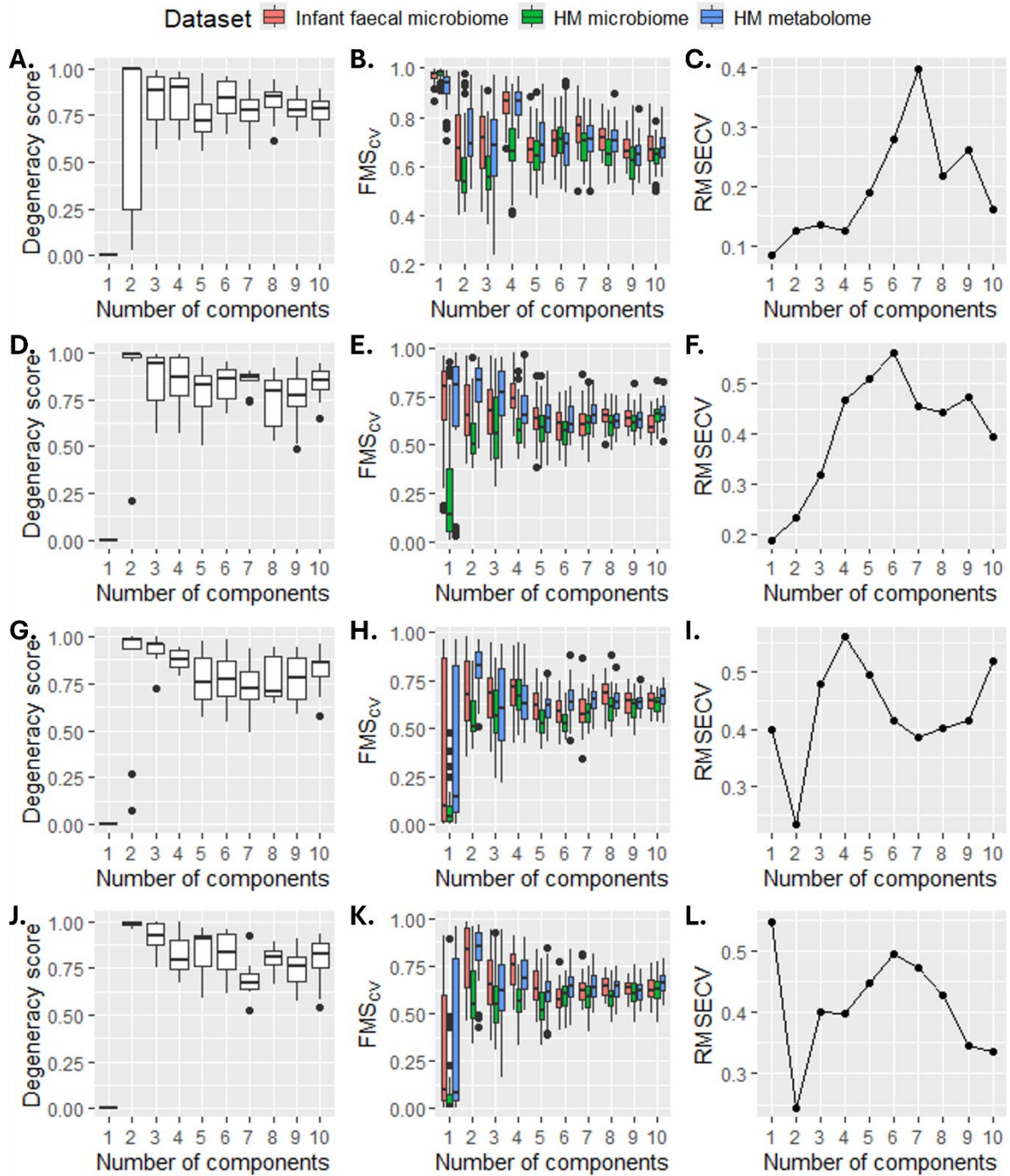
167



168

169 **Supplementary Figure 11. Factor recovery in ACMTF and ACMTF-R for varying noise levels on y .** One
 170 hundred randomly initialised eight-component ACMTF-R and ACMTF models were fitted for each
 171 combination of noise on y (η_y) and tuning parameter (π), after which the lowest-loss models were
 172 compared. Outcome per setting for (A) overall factor recovery, (B) factor recovery of the hidden
 173 component, (C) shared subject mode recovery of the hidden component, (D) feature mode recovery
 174 of $\underline{X}^{(1)}$ for the hidden component, (E) time mode recovery of $\underline{X}^{(1)}$ for the hidden component, (F)
 175 feature mode recovery of $\underline{X}^{(3)}$ for the hidden component, (G) time mode recovery of $\underline{X}^{(3)}$ for the
 176 hidden component.

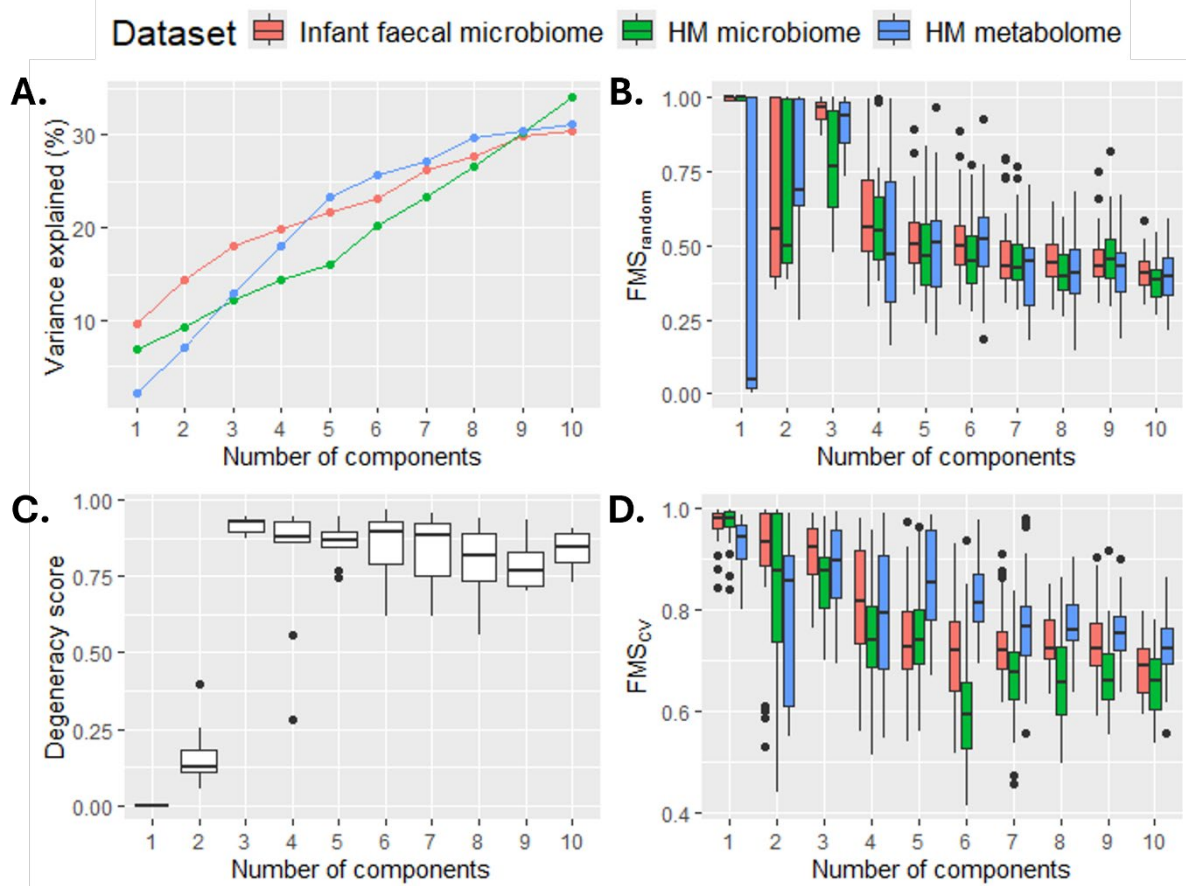
177



178

179 **Supplementary Figure 12.** Model selection to determine the correct number of components for
 180 ACMTF-R regressing on maternal ppBMI at (A-C) $\pi = 0.80$, (D-F) $\pi = 0.85$, (G-I) $\pi = 0.90$, (J-L) $\pi =$
 181 0.95.

182



183

184 **Supplementary Figure 13.** Model selection outcome for the biological application using ACMTF to
 185 determine the appropriate number of components. (A) Variance explained per block. (B) FMS_{random}
 186 across 10 randomly initialised models per setting. (C) Degeneracy score across the (same) randomly
 187 initialised models. (D) FMS_{cv} across 10 CV folds.

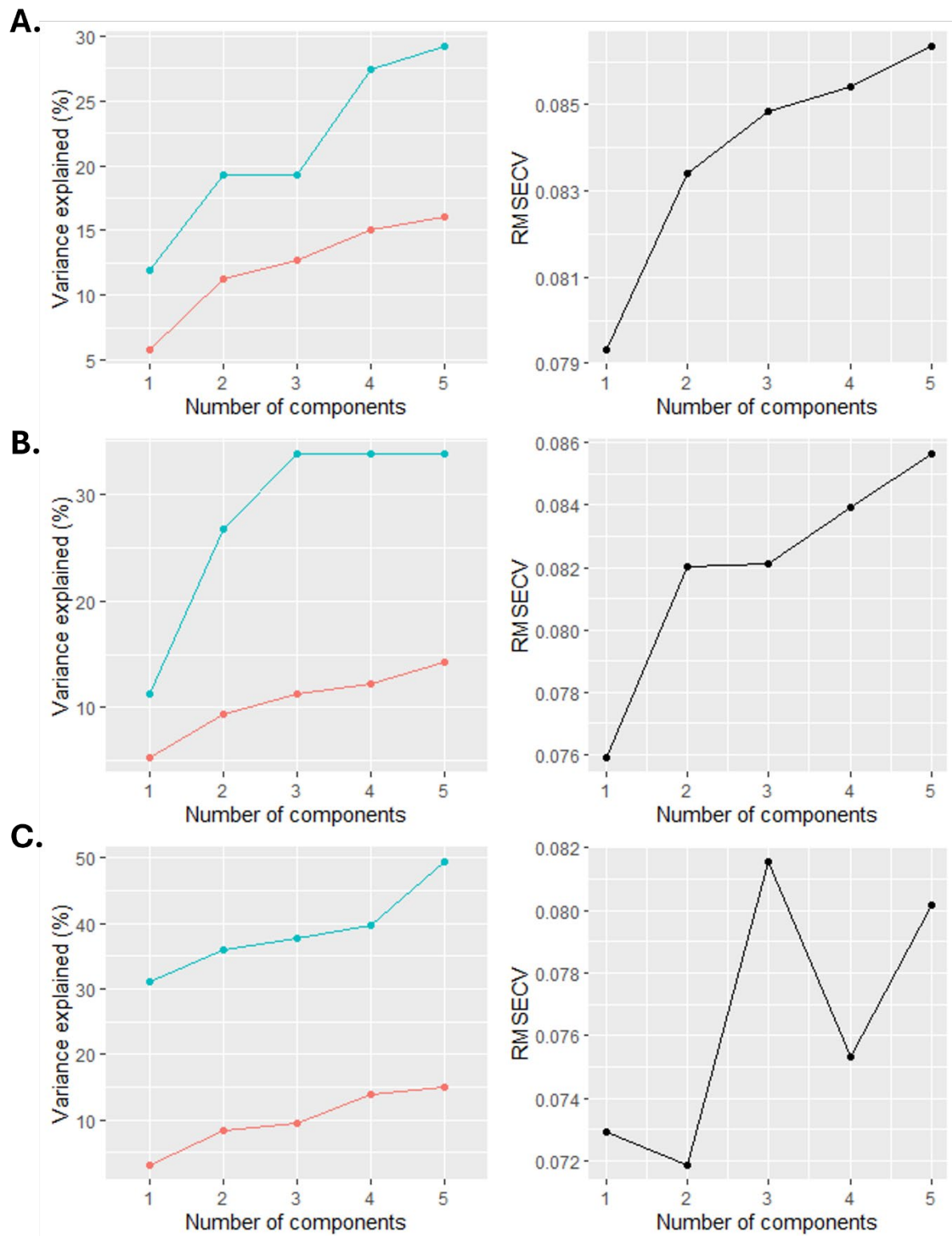
188

189

190

191

192



193

194 **Supplementary Figure 14.** Cross-validation outcome for NPLS for the (A) infant faecal microbiome,

195 (B) HM milk, (C) HM metabolomics datasets of the biological application.

196

197 **Supplementary Tables**

198 **Supplementary Table 1.** Results of a Kendall's tau test for the feature mode loadings of the lowest-
 199 loss ACMTF-R models against the true feature mode loadings per data block for the hidden local
 200 component L_{13} . Values correspond to τ and stars indicate their significance (*: $p < 0.05$, **: $p <$
 201 0.01 , ***: $p < 0.001$).

π	$\underline{X}^{(1)}$		$\underline{X}^{(3)}$	
	τ	95% CI	τ	95% CI
1.00 (ACMTF)	0.00	-0.10 – 0.09	-0.04	-0.13 – 0.06
0.90	0.44***	0.37 – 0.51	0.36***	0.27 – 0.45
0.80	0.45***	0.38 – 0.51	0.36***	0.27 – 0.45
0.70	0.45***	0.38 – 0.52	0.37***	0.28 – 0.45
0.60	0.45***	0.38 – 0.52	0.36***	0.28 – 0.44
0.50	0.43***	0.35 – 0.50	0.35***	0.26 – 0.43
0.40	0.43***	0.35 – 0.50	0.35***	0.26 – 0.44
0.30	0.03	0.08 – 0.12	0.00	-0.10 – 0.10
0.20	0.09	0.00 – 0.18	-0.05	-0.15 – 0.04
0.10	0.02	-0.08 – 0.11	-0.06	-0.15 – 0.03

202

203

204 **Supplementary Table 2.** Correlation matrix of the MAINHEALTH metadata. Birth mode, secretor
 205 status (*Se*), and Lewis status (*Le*) are binarized variables.

	ppBMI	Birth mode	<i>Se</i>	<i>Le</i>	WHZ
ppBMI	1	0.04	-0.02	0.06	0.07
Birth mode		1	-0.06	-0.05	0.03
<i>Se</i>			1	-0.10	-0.08
<i>Le</i>				1	-0.18
WHZ					1

206