Some new methods for three-mode factor analysis and multi-set factor analysis

Lam, Tam
Chapter 5

Three-mode factor analysis by means of Tucker3

5.1 Introduction

In this chapter we discuss a different approach to three-mode factor analysis that is based on the Tucker3 model of [Tucker (1966)]. Unlike CP, in the Tucker3 model, each mode of the data may have a different number of components, and all their interaction strengths are included as entries $g_{rpq}$ of the so called core array. Suppose that we have $R$ components for the $N$ observations, $P$ components for the $J$ variables, and $Q$ components for the $K$ conditions. The Tucker3 model can then be written in the form of

$$
X_{(N \times JK)} = FG(C \otimes B)^T + E_{(N \times JK)}, \quad (5.1)
$$

103
with matrix $G$ ($R \times PQ$) of interaction strengths as follows.

$$G = \begin{bmatrix}
g_{111} & \cdots & g_{1P1} & \cdots & g_{11Q} & \cdots & g_{1PQ} \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
g_{R11} & \cdots & g_{RP1} & \cdots & g_{RIQ} & \cdots & g_{RPQ}
\end{bmatrix}.$$ 

It is well known that the three-mode Tucker model is not unique. All of $F$, $C$, and $B$ can be rotated, with inverse transformations applied to the interaction strengths in $G$, without affecting the model part $FG(C \otimes B)^T$; see section 2.2.6. The covariance model corresponding to (5.1) is given as

$$\Sigma = (C \otimes B) G^T \Phi G (C \otimes B)^T + U,$$  

where $C$ is an $K \times Q$ matrix containing the $Q$ method components as columns, $B$ is a $J \times P$ matrix containing the $P$ variable components as columns, $C \otimes B$ is the Kronecker product of $C$ and $B$.

For simplicity, one may consider $\Psi = G^T \Phi G$ as factor covariance matrix. The case of diagonal $\Psi$ can be rewritten as $\Sigma = (CC^T \otimes BB^T) + U$ and is known as the direct product model; see Browne (1984) or Wothke (1996). In this model, $CC^T$ and $BB^T$ may be seen as covariance matrices corresponding to methods and variables, respectively.

Fitting the covariance model (5.2) can be done by directly fitting the component model (5.1) to the data. Alternating least squares algorithms minimizing the sum-of-squares of $E_{(N \times JK)}$ can be found in Kroonenberg and De Leeuw (1980) and Kiers et al. (1992), see section 2.2.5. Contrary to the confirmatory factor analysis approach, convergence problems do not often occur. One may also fit the covariance model (5.2) by an algorithm for nonlinear optimization. For example, Bentler and Lee (1978, 1979) propose to use a Gauss-Newton algorithm. For an overview of three-mode component and factor models based
on [Tucker (1966), we refer to [Kroonenberg and Oort (2003). For an accessible introduction to three-mode component analysis, see Kiers and Van Mechelen (2001).]

In section 5.2, we provide an estimation procedure to fit model (5.2) such that $\Sigma - U$ is a covariance matrix. This allows for the computation of the percentage of explained common variance. The estimation procedure is analogous to section 4.2 for three-mode factor analysis by means of CP. In sections 5.3 and 5.4, we apply model (5.2) to datasets in the literature. Finally, section 5.5 contains a discussion of our findings.

### 5.2 Estimation Procedure

Here, we present our estimation procedure for the Tucker3 covariance model (5.2). After having computed the data covariance matrix $\Sigma$, the steps of our estimation procedure for model (5.2) are as follows.

1. Use the MRFA algorithm of [Ten Berge and Kiers (1991)] to estimate $U$. This implies that $U$ is nonnegative, $\Sigma - U$ is a covariance matrix, and the trace of $\Sigma - L - U$ is minimal, where $L$ is a best rank-$PQ$ approximation of $\Sigma - U$.

2. Compute the eigendecomposition $\Sigma - U = VSV^T$, with $V$ having orthonormal columns, and $S$ the diagonal matrix containing the eigenvalues in decreasing order. This is also the singular value decomposition of $\Sigma - U$. Let $P = VS^{1/2}$, which implies $\Sigma - U = PP^T$.

3. Fit the Tucker3 model as $P \approx (C \otimes B)G^T T^T$ by using the alternating least squares algorithm of [Kroonenberg and De Leeuw (1980)] with number of components in each mode respectively $R, P, Q$. Matrix $P$ ($JK \times JK$) is
the matricized \( K \times J \times JK \) array, \( G (R \times PQ) \) is the matricized \( R \times P \times Q \) core array, and \( T \) is a \( JK \times R \) matrix satisfying \( T^T T = I_R \). The columns of \( B, C, \) and \( T \) are scaled such that they have length 1. We obtain \( \Sigma - U \approx (C \otimes B)\Psi(K \otimes B)^T \), where \( \Psi = G^T T^T TG = G^T G \) is an interaction matrix of the factors in the second and third modes. We set \( R = PQ \) to not constrain the rank of \( \Psi \). We use Joint Orthomax rotation of (Kiers, 1998a) to rotate the Tucker3 solution to simple structure in \( B, C, \) and in the core \( G \). We use the standard weights specified in Kiers (1998a); see also section 3.6.2. We evaluate the Tucker3 fit as

\[
100 - 100 \cdot \frac{\text{ssq}(P - (C \otimes B)G^T T^T)}{\text{ssq}(P)},
\]

which is the percentage of the sum-of-squares of \( P \) that is fitted by \((C \otimes B)G^T T^T\). In the alternating least squares algorithm, \((C \otimes B)G^T T^T\) is the regression of \( P \) on \((C \otimes B)G^T\). Since the regression and the residual are orthogonal, it follows that (5.3) is equal to

\[
100 \cdot \frac{\text{ssq}((C \otimes B)G^T T^T)}{\text{ssq}(P)}.
\]

Since we were not able to construct an algorithm for the simultaneous estimation of \( U \) and \( C, B, \Psi \) under the restriction that \( \Sigma - U \) is a covariance matrix, we instead estimate \( U \) and \( C, B, \Psi \) sequentially. First, we estimate \( U \) by MRFA based on a rank-\( PQ \) factor model for \( \Sigma - U \). For fitting \((C \otimes B)\Psi(K \otimes B)^T\) to \( \Sigma - U \), we have chosen to fit Tucker3 to \( P \) with \( \Sigma - U = PP^T \). We run the algorithm of Tucker3 that is programmed by Henk A. L. Kiers 10 times for random starting with convergence criterion \( 1e - 9 \) and keep the solution with the highest fit percentage. The percentage of explained common variance in the Tucker3 covariance model is given by (5.3), which can be written as

\[
100 \cdot \frac{\text{trace}((C \otimes B)\Psi(K \otimes B)^T)}{\text{trace}(\Sigma - U)}.
\]
To obtain a percentage of explained common variance for each variable-condition combination separately, we proceed as follows. Contrary to the two-mode case of MRFA, it may happen that some diagonal entry of \((C \otimes B)\Psi(C \otimes B)^T\) is larger than the corresponding communality on the diagonal of \(\Sigma - U\). Because of this, we formulate the explained common variance per variable-condition combination analogous to (5.3) rather than to (5.4). A particular variable-condition corresponds to a row of \(P - (C \otimes B)G^T T^T\). Let row \(m\) of this matrix be denoted as \(q_m^T\). Then we define the corresponding percentage of explained common variance as

\[
100 - 100 \cdot \frac{ssq(q_m^T)}{(\Sigma - U)_{mm}},
\]

where \((\Sigma - U)_{mm}\) is the corresponding communality.

As mentioned in section 4.1.1, \((C \otimes B)\Psi\) is also the covariance matrix between the variables and factors. This matrix is used to interpret the factors.

### 5.3 Application I

In this section, we apply our model to a data set from [Dickinson and Tice (1973)](Dickinson and Tice (1973)). The correlation matrix of the [Dickinson and Tice (1973)](Dickinson and Tice (1973)) job behavior ratings is reproduced in Table 5.1. This data set is obtained from \(N = 149\) subjects with \(J = 3\) traits assessed by \(K = 3\) methods. The three traits are: Getting along with others (G), Dedication (D), and Ability to Apply Learning (L). And the three methods are: Peer Nominations (PN), Peer Checklist ratings (PC), and Supervisor Checklist ratings (SC).

In section 5.3.1 we fit our three-mode CP factor model (4.9) to this data set. In section 5.3.2 we fit our three-mode Tucker3 factor model (5.2) to this data set and compare the results to those in section 5.3.1.
5.3.1 Three-mode CP factor analysis solution

Here, we fit our three-mode CP factor model (4.9) to this data set. Following is the three-mode CP factor solution with $R = 2$ orthogonal factors

$$B = \begin{pmatrix} 0.64 & 0.74 \\ 0.62 & 0.58 \\ 0.45 & 0.34 \end{pmatrix}, \quad C = \begin{pmatrix} 0.99 & 0.68 \\ 0.59 & -0.64 \\ 0.74 & -0.61 \end{pmatrix}$$

and the unique variance $U = \text{diag}(0.18, 0.38, 0.59, 0.49, 0.47, 0.67, 0.15, 0.55, 0.68)$. Matrix $B$ is rescaled to have columns of length 1. The percentage of explained common variance is 64.68%, where 38.81% is due to factor 1, and 25.87% is due to factor 2. The percentages of explained common variance for each trait-

Table 5.1: Correlations of three traits Getting along with others (G), Dedication (D), and Ability to Apply Learning (L) measured by three methods Peer Nominations (PN), Peer Checklist ratings (PC), and Supervisor Checklist ratings (SC).

<table>
<thead>
<tr>
<th></th>
<th>PN</th>
<th></th>
<th>PC</th>
<th></th>
<th>SC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
<td>D</td>
<td>L</td>
<td>G</td>
<td>D</td>
<td>L</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0.524</td>
<td>0.241</td>
<td>0.071</td>
<td>0.022</td>
<td>0.076</td>
</tr>
<tr>
<td>D</td>
<td>0.524</td>
<td>1</td>
<td>0.403</td>
<td>0.102</td>
<td>0.096</td>
<td>0.102</td>
</tr>
<tr>
<td>L</td>
<td>0.241</td>
<td>0.403</td>
<td>1</td>
<td>-0.018</td>
<td>0.018</td>
<td>0.100</td>
</tr>
<tr>
<td>G</td>
<td>0.071</td>
<td>0.102</td>
<td>-0.018</td>
<td>1</td>
<td>0.435</td>
<td>0.342</td>
</tr>
<tr>
<td>D</td>
<td>0.022</td>
<td>0.096</td>
<td>0.018</td>
<td>0.435</td>
<td>1</td>
<td>0.347</td>
</tr>
<tr>
<td>L</td>
<td>0.076</td>
<td>0.102</td>
<td>0.100</td>
<td>0.342</td>
<td>0.347</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>0.136</td>
<td>0.132</td>
<td>0.061</td>
<td>0.243</td>
<td>0.093</td>
<td>0.053</td>
</tr>
<tr>
<td>D</td>
<td>-0.028</td>
<td>0.168</td>
<td>0.135</td>
<td>0.093</td>
<td>0.209</td>
<td>0.108</td>
</tr>
<tr>
<td>L</td>
<td>-0.054</td>
<td>0.162</td>
<td>0.252</td>
<td>0.053</td>
<td>0.108</td>
<td>0.108</td>
</tr>
</tbody>
</table>

Note: Data taken from Dickinson and Tice (1973)
method combination are given in column 8 of Table 5.4. They are between 44 and 91 percent, where the percentages of L-PC and L-SC are rather low (44% and 45%). The model has 10 parameters, not counting unique variances.

For interpretation of (5.5) we compute \((C \odot B)\Phi\); see columns 4 and 5 of Table 5.2. Factor 1 is a general factor with larger loadings for Peer Nominations (PN) and Supervisor Checklist ratings (SC). Factor 2 represents a contrast between Peer Nominations (PN) and Peer Checklist ratings (PC) and Supervisor Checklist ratings (SC), mostly for Getting along with others (G) and Dedication (D).

The solution with \(R = 3\) orthogonal factors is:

\[
B = \begin{pmatrix}
0.63 & 0.79 & 0.61 \\
0.63 & 0.52 & 0.61 \\
0.46 & 0.30 & 0.49 \\
\end{pmatrix}
\]

\[
C = \begin{pmatrix}
1.04 & -0.44 & -0.41 \\
0.62 & 0.20 & 0.90 \\
0.61 & 0.97 & -0.17 \\
\end{pmatrix}
\]

and the unique variance \(U = \text{diag}(0.24, 0.36, 0.60, 0.49, 0.47, 0.67, 0.08, 0.56, 0.67)\).

Matrix \(B\) is rescaled to have columns of length 1. The percentage of explained common variance is 82.76%, where 37.60% is due to factor 1, 24.47% is due to factor 2, and 20.69% is due to factor 3. The percentages of explained common variance for each trait-method combination are given in column 10 of Table 5.4. They are between 59 and 94 percent, where the percentages of L-PC and L-SC are much larger than those in case \(R = 2\). The model has 15 parameters.

For interpretation of (5.6) we compute \((C \odot B)\Phi\); see columns 9, 10, 11 of Table 5.2. Each factor loads high on one method. Factor 1 is a general factor with highest loadings for Peer Nominations (PN). Factor 2 can be interpreted as a combination between Getting along with others (G) and Dedication (D) for Supervisor Checklist ratings (SC). Factor 3 is also a general factor for Peer Checklist ratings (PC).
We also obtain similar matrices $B$ in the case of oblique factors with $R = 2$ and $R = 3$. The oblique solutions yield highly correlated factors (0.69 for $R = 2$ and 0.84 for $R = 3$). This implies that the solutions with oblique factors give less clear interpretation than with orthogonal factors; see Table 5.2.

Table 5.2: Values of $(C \odot B)\Phi$ for the estimated CP covariance model \((4.9)\) with $R = 2$ and $R = 3$ factors fitted to the MTMM data from Dickinson and Tice (1973).

<table>
<thead>
<tr>
<th>variable</th>
<th>$R = 2$ oblique</th>
<th>orthogonal</th>
<th>$R = 3$ oblique</th>
<th>orthogonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-PN</td>
<td>0.33 -0.32</td>
<td>0.64 0.50</td>
<td>0.38 0.14 -0.14</td>
<td>0.65 -0.35 -0.25</td>
</tr>
<tr>
<td>D-PN</td>
<td>0.44 -0.12</td>
<td>0.61 0.40</td>
<td>0.46 0.16 0.07</td>
<td>0.65 -0.23 -0.25</td>
</tr>
<tr>
<td>L-PN</td>
<td>0.32 -0.05</td>
<td>0.44 0.23</td>
<td>0.28 0.04 0.07</td>
<td>0.47 -0.13 -0.20</td>
</tr>
<tr>
<td>G-PC</td>
<td>0.54 0.55</td>
<td>0.37 -0.46</td>
<td>0.59 0.68 0.28</td>
<td>0.38 0.16 0.55</td>
</tr>
<tr>
<td>D-PC</td>
<td>0.48 0.46</td>
<td>0.36 -0.37</td>
<td>0.58 0.68 0.25</td>
<td>0.38 0.10 0.55</td>
</tr>
<tr>
<td>L-PC</td>
<td>0.32 0.31</td>
<td>0.26 -0.21</td>
<td>0.46 0.54 0.19</td>
<td>0.28 0.06 0.44</td>
</tr>
<tr>
<td>G-SC</td>
<td>0.63 0.58</td>
<td>0.47 -0.45</td>
<td>0.54 0.23 0.85</td>
<td>0.37 0.77 -0.10</td>
</tr>
<tr>
<td>D-SC</td>
<td>0.57 0.50</td>
<td>0.45 -0.35</td>
<td>0.40 0.15 0.62</td>
<td>0.38 0.51 -0.10</td>
</tr>
<tr>
<td>L-SC</td>
<td>0.38 0.33</td>
<td>0.33 -0.21</td>
<td>0.24 0.07 0.39</td>
<td>0.27 0.29 -0.08</td>
</tr>
</tbody>
</table>

*Note:* Numbers larger than 0.4 are in boldfont.

### 5.3.2 Three-mode Tucker3 factor analysis solution

Next, we fit our model \((5.2)\) to the data set of Dickinson and Tice (1973). The explained common variances for $P \in \{1, 2\}$, and $Q \in \{2, 3\}$ are given in the following table.
Below, we present the solutions with $P = 1, P = 2, P = Q = 2, \text{ and } P = 1, Q = 3$. And we compare these solutions to CP solutions with $R = PQ$.

After rotating, the solution with $P = 1$ and $Q = 2$ is as follows:

$$
\mathbf{B} = \begin{pmatrix} 0.68 \\ 0.60 \\ 0.41 \end{pmatrix}, \quad \mathbf{G} = \mathbf{D}, \quad \mathbf{C} = \begin{pmatrix} 0.01 & 0.99 \\ 0.70 & -0.09 \\ 0.71 & 0.06 \end{pmatrix}, \quad \mathbf{L}, \quad \mathbf{D}, \quad \mathbf{P}, \quad \mathbf{N}
$$

$$
\mathbf{G} = \begin{pmatrix} -1.29 & -0.14 \\ 0.12 & 1.19 \end{pmatrix}, \quad \Psi = \begin{pmatrix} 1.68 & 0.33 \\ 0.33 & 1.44 \end{pmatrix}.
$$

The unique variance is: $\text{diag}(\mathbf{U}) = (0.18 \ 0.38 \ 0.59 \ 0.48 \ 0.47 \ 0.67 \ 0.15 \ 0.55 \ 0.68)$. The percentage of explained common variance is: 64.41. The percentages of explained common variance for each trait-method combination are given in column 2 of Table 5.4. They are between 41 and 90 percent, where the percentages of L-PC and L-SC are rather low (45% and 41%). Since both matrices $\mathbf{B}$ and $\mathbf{C}$ are rescaled to have columns of length 1, $\mathbf{C}$ is very close to the following form

$$
\begin{pmatrix} 0 \\ 1 \\ * \ 0 \\ * \ 0 \end{pmatrix},
$$

and $\Psi$ is symmetric, it follows that we obtain 2 parameters from $\mathbf{B}$, 1 parameter from $\mathbf{C}$, and 3 parameters from $\Psi$. Hence, the model has 6 parameters in total.
For interpretation of (5.7) we also compute \((C \otimes B)\Psi\); see columns 2,3 of Table 5.3. The solution (5.7) can be interpreted as follows. Only entries on the diagonal of \(\Psi\) are large and each diagonal entry of \(\Psi\) corresponds to a column of \(C \otimes B = [c_1 \otimes B | c_2 \otimes B]\). Factor 1 is interpreted as a general factor but only for Peer Checklist ratings (PC) and Supervisor Checklist ratings (SC). Factor 2 is also a general factor but only for Peer Nominations (PN). Both factors have smaller loadings for Ability to Apply Learning (L).

Compared to the solution (5.5) using CP, matrix \(C\) in (5.7) has a clearer interpretation than \(C\) in (5.5). Matrix \(B\) in (5.7) has only one column which is very similar to the first column of \(B\) in (5.5). Since the Tucker3 factor model has less parameters than those of the CP factor model, solution (5.7) is more parsimonious than solution (5.5).

After rotating, the solution with \(P = 2\) and \(Q = 2\) is as follows:

\[
B = \begin{pmatrix}
0.70 & -0.65 \\
0.60 & 0.35 \\
0.38 & 0.66
\end{pmatrix}, \quad C = \begin{pmatrix}
0.14 & 0.97 \\
0.61 & -0.23 \\
0.78 & 0.00
\end{pmatrix}, \quad G = \begin{pmatrix}
0.08 & 0.03 & 1.13 & 0.06 \\
-1.33 & 0.00 & -0.12 & -0.10 \\
-0.01 & -0.59 & -0.01 & -0.37 \\
-0.01 & -0.01 & -0.00 & -0.31
\end{pmatrix}, \quad \Psi = \begin{pmatrix}
1.79 & 0.00 & 0.26 & 0.14 \\
0.00 & 0.35 & 0.05 & 0.22 \\
0.26 & 0.05 & 1.30 & 0.08 \\
0.14 & 0.22 & 0.08 & 0.25
\end{pmatrix}.
\]

(5.8)

The unique variance is: \(\text{diag}(U) = (0.30 \ 0.34 \ 0.60 \ 0.50 \ 0.47 \ 0.66 \ 0.00 \ 0.56 \ 0.67)\). The percentage of explained common variance is: 76.11. The percentages of explained common variance for each trait-method combination are given in column 4 of Table 5.4. They are between 46 and 98 percent, where the percentage of G-PC is rather low (46%). Since both \(B\) and \(C\) are rescaled to have columns
of length 1, \( C \) has one entry that equals to zero, and \( \Psi \) is symmetric with one zero entry, this implies that \( B \) has 4 parameters, \( C \) has 3 parameters, and \( \Psi \) has 9 parameters. Hence, the model has 16 parameters in total.

For interpretation of (5.8) we also compute \((C \otimes B)\Psi\); see columns 4, 5, 6, 7 of Table 5.3. The solution (5.8) can be interpreted as follows. Only the first and third entries on the diagonal of \( \Psi \) are large and each diagonal entry of \( \Psi \) corresponds to a column of \( C \otimes B = [c_1 \otimes b_1 | c_1 \otimes b_2 | c_2 \otimes b_1 | c_2 \otimes b_2] \). This implies that columns \( c_1 \otimes b_2 \) and \( c_2 \otimes b_2 \) play only a minor role in the interpretation. Hence, this solution is interpreted almost the same to the interpretation of the solution (5.7).

After rotating, the result with \( P = 1 \) and \( Q = 3 \) is:

\[
B = \begin{pmatrix}
0.68 \\
0.60 \\
0.42
\end{pmatrix}
G
D,
L
C = \begin{pmatrix}
1.00 & -0.01 & -0.01 \\
0.01 & 0.01 & 1.00 \\
0.01 & 1.00 & -0.01
\end{pmatrix}
PN
PC,
SC
G = \begin{pmatrix}
1.18 & 0.12 & 0.07 \\
-0.06 & -0.16 & -1.07 \\
-0.11 & -1.13 & -0.18
\end{pmatrix}
,\quad \Psi = \begin{pmatrix}
1.42 & 0.29 & 0.17 \\
0.29 & 1.33 & 0.40 \\
0.17 & 0.40 & 1.19
\end{pmatrix}
(5.9)
\]

The unique variance is: \( \text{diag}(U) = (0.24 \ 0.36 \ 0.60 \ 0.49 \ 0.47 \ 0.67 \ 0.08 \ 0.56 \ 0.67) \).

The percentage of explained common variance is: 81.69. The explained common variances for all trait-method combinations are given in column 6 of Table 5.4. They are between 58 and 91 percent. Most trait-method combinations with low percentages of explained common variance in the solutions with \( P = 1, Q = 2 \) and \( P = Q = 2 \) have much larger percentage of explained common variance in the \( P = 1, Q = 3 \) solution. Since both \( B \) and \( C \) are rescaled to have columns of length 1, \( C \) is almost completely fixed, and \( \Psi \) is symmetric, this follows that we obtain 2 parameters from \( B \), and 6 parameters from \( \Psi \). Therefore, the model
has 8 parameters in total.
For interpretation of (5.9) we also compute \((C \otimes B)\Psi\); see columns 8, 9, 10 of Table 5.3. The solution (5.9) can be interpreted as follows. Since only entries on the diagonal of \(\Psi\) are large and each diagonal entry of \(\Psi\) corresponds to a column of \(C \otimes B = [c_1 \otimes B | c_2 \otimes B | c_3 \otimes B]\), this implies that we obtain one method per factor. Factor 1 is a general factor but only for Supervisor Checklist ratings (SC). Factor 2 is a general factor but only for Peer Nominations (PN). Factor 3 is a general factor but only for Peer Checklist ratings (PC). All factors having smaller loadings for Ability to Apply Learning (L).
Compared to the solutions (5.5) and (5.6) using CP, matrices \(C\) in the Tucker3 solutions (5.7) and (5.9) have much more weights that are very close to zero. And each column of \(C\) in Tucker3 solution has at least one entry close to zero and at least one entry close to 1. This means that matrices \(C\) in Tucker3 solutions have a clearer interpretation than \(C\) in CP solutions. Matrices \(B\) in (5.7) and (5.9) have only one column instead of two similar columns of \(B\) in (5.5) and (5.6).
Since Tucker3 solutions have less parameters than the CP solutions with \(R = PQ\) factors, the Tucker3 solutions are more parsimonious than the corresponding CP solutions. However, the percentages of explained common variance of Tucker3 are a little bit less than of CP solutions with \(R = PQ\).
Table 5.3: Values of \((C \otimes B)\Psi\) for the estimated Tucker3 covariance model (5.2) with \(P\) traits factors and \(Q\) method factors, fitted to the MTMM data from Dickinson and Tice (1973).

<table>
<thead>
<tr>
<th>variable</th>
<th>(P = 1, Q = 2)</th>
<th>(P = 2, Q = 2)</th>
<th>(P = 1, Q = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-PN</td>
<td>0.24 <strong>0.98</strong></td>
<td>0.26 -0.14 <strong>0.86</strong> -0.11</td>
<td><strong>0.97</strong> 0.18 0.10</td>
</tr>
<tr>
<td>D-PN</td>
<td>0.22 <strong>0.87</strong></td>
<td>0.35 0.12 <strong>0.81</strong> 0.16</td>
<td><strong>0.85</strong> 0.16 0.09</td>
</tr>
<tr>
<td>L-PN</td>
<td>0.14 <strong>0.59</strong></td>
<td>0.29 0.19 <strong>0.56</strong> 0.22</td>
<td><strong>0.59</strong> 0.11 0.06</td>
</tr>
<tr>
<td>G-PC</td>
<td><strong>0.78</strong> 0.06</td>
<td><strong>0.75</strong> -0.11 -0.10 -0.00</td>
<td>0.13 0.29 <strong>0.82</strong></td>
</tr>
<tr>
<td>D-PC</td>
<td><strong>0.69</strong> 0.05</td>
<td><strong>0.61</strong> 0.05 -0.07 0.06</td>
<td>0.12 0.25 <strong>0.72</strong></td>
</tr>
<tr>
<td>L-PC</td>
<td><strong>0.47</strong> 0.03</td>
<td>0.37 0.10 -0.04 0.07</td>
<td>0.08 0.18 <strong>0.50</strong></td>
</tr>
<tr>
<td>G-SC</td>
<td><strong>0.83</strong> 0.22</td>
<td><strong>0.98</strong> -0.18 0.12 -0.03</td>
<td>0.21 <strong>0.91</strong> 0.26</td>
</tr>
<tr>
<td>D-SC</td>
<td><strong>0.74</strong> 0.20</td>
<td><strong>0.84</strong> 0.09 0.14 0.13</td>
<td>0.18 <strong>0.79</strong> 0.22</td>
</tr>
<tr>
<td>L-SC</td>
<td><strong>0.50</strong> 0.13</td>
<td><strong>0.54</strong> 0.18 0.10 0.16</td>
<td>0.12 <strong>0.56</strong> 0.15</td>
</tr>
</tbody>
</table>

*Note:* Numbers larger than 0.4 are in boldfont.
Table 5.4: The explained common variances for all trait-method combinations and the communalities for the MTMM data from Dickinson and Tice (1973)

<table>
<thead>
<tr>
<th>variable</th>
<th>CP with ( R = 3 )</th>
<th>CP with ( R = 2 )</th>
<th>( P = 1, Q = 2 )</th>
<th>( P = 2, Q = 2 )</th>
<th>( P = 1, Q = 3 )</th>
<th>( P = 2, Q = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
</tr>
<tr>
<td>G-PN</td>
<td>78.82</td>
<td>79.81</td>
<td>98.91</td>
<td>93.93</td>
<td>91.91</td>
<td>91.91</td>
</tr>
<tr>
<td>D-PN</td>
<td>90.61</td>
<td>93.63</td>
<td>91.66</td>
<td>66.56</td>
<td>63.56</td>
<td>63.56</td>
</tr>
<tr>
<td>L-PN</td>
<td>58.33</td>
<td>54.33</td>
<td>58.33</td>
<td>55.35</td>
<td>55.35</td>
<td>55.35</td>
</tr>
<tr>
<td>G-PC</td>
<td>58.51</td>
<td>53.53</td>
<td>91.50</td>
<td>53.90</td>
<td>53.90</td>
<td>53.90</td>
</tr>
<tr>
<td>D-PC</td>
<td>65.50</td>
<td>66.66</td>
<td>91.50</td>
<td>53.90</td>
<td>53.90</td>
<td>53.90</td>
</tr>
<tr>
<td>L-PC</td>
<td>45.33</td>
<td>45.33</td>
<td>45.33</td>
<td>45.33</td>
<td>45.33</td>
<td>45.33</td>
</tr>
<tr>
<td>G-SC</td>
<td>61.85</td>
<td>61.85</td>
<td>61.85</td>
<td>61.85</td>
<td>61.85</td>
<td>61.85</td>
</tr>
<tr>
<td>D-SC</td>
<td>55.44</td>
<td>54.45</td>
<td>55.45</td>
<td>55.45</td>
<td>55.45</td>
<td>55.45</td>
</tr>
<tr>
<td>L-SC</td>
<td>41.32</td>
<td>41.32</td>
<td>41.32</td>
<td>41.32</td>
<td>41.32</td>
<td>41.32</td>
</tr>
</tbody>
</table>

ECV% total:
- 64.41
- 76.11
- 81.69

81.69
5.4 Application II

In this section, we apply our three-mode Tucker\(3\) factor model (5.2) to BJW data from section 4.3.2 for which we fitted three-mode CP factor model (4.9). To compare to the CP solutions in section 4.3.2 we fit three-mode Tucker\(3\) factor model (5.2) to this dataset with \(P \in \{2, 3\}\) and \(Q \in \{1, 2\}\). The explained common variances for \(P \in \{2, 3\}\), and \(Q \in \{1, 2\}\) are given in the following table.

<table>
<thead>
<tr>
<th>P</th>
<th>Q</th>
<th>ECV% for Tucker(3)</th>
<th>ECV% for CP with (R = PQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>71.00</td>
<td>74.51</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>74.76</td>
<td>84.54</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>86.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>91.52</td>
<td></td>
</tr>
</tbody>
</table>

Below, we present the solutions with \(P = 2, Q = 1\) and \(P = 3, Q = 1\).

After rotating, the result with \(P = 2\) and \(Q = 1\) is:

\[
B = \begin{pmatrix}
0.33 & 0.13 & \text{item 1} \\
0.41 & -0.17 & \text{item 2} \\
0.41 & -0.25 & \text{item 3} \\
0.33 & 0.28 & \text{item 4} \\
0.10 & \textbf{0.78} & \text{item 5} \\
0.33 & 0.31 & \text{item 6} \\
0.38 & -0.25 & \text{item 7} \\
0.40 & -0.14 & \text{item 8}
\end{pmatrix}, \quad 
C = \begin{pmatrix}
0.70 \\
0.72
\end{pmatrix}, \quad \text{Yourself},

\[
\Psi = \begin{pmatrix}
7.61 & 1.26 \\
1.26 & 1.73
\end{pmatrix}, \quad \text{Others},
\]

\[
G = \begin{pmatrix}
\textbf{-2.76} & -0.42 \\
0.07 & \textbf{1.24}
\end{pmatrix}, \quad \Psi = \begin{pmatrix}
7.61 & 1.26 \\
1.26 & 1.73
\end{pmatrix}. \quad (5.10)
\]
The unique variance is: \( \text{diag}(U) = (0.34 \ 0.12 \ 0.16 \ 0.19 \ 0.30 \ 0.15 \ 0.21 \ 0.11 \ 0.21 \ 0.24 \ 0.15 \ 0.23 \ 0.00 \ 0.11 \ 0.20 \ 0.12) \). The percentage of explained common variance is: 71.00. The explained common variances for all item-condition combinations are given in column 2 of Table 5.6. They are between 58 and 85 percent with low percentage for items 7 and 8 for Others (58%).

For interpreting (5.10) we compute \((C \otimes B)\Psi\); see columns 2, 3 of Table 5.10.

The solution (5.10) can be interpreted as follows. Factor 1 is a general factor. Factor 2 is a combination of items 4, 5, and 6. There is not much difference between two conditions.

Compared to the CP solution with \(R = PQ = 2\), this solution has less percent explained common variance than of the CP solution (71.00\% versus 74.51\%). In CP solution, there is one general factor and one "for Yourself" factor.

After rotating, the result with \(P = 3\) and \(Q = 1\) is:

\[
B = \begin{pmatrix}
0.32 & -0.17 & 0.72 \\
0.40 & -0.25 & 0.09 \\
0.39 & -0.17 & -0.35 \\
0.35 & 0.17 & 0.24 \\
0.18 & 0.87 & -0.08 \\
0.36 & 0.23 & 0.17 \\
0.36 & -0.17 & -0.33 \\
0.40 & -0.04 & -0.36
\end{pmatrix}
\]

\[
G = \begin{pmatrix}
-2.79 & -0.20 & -0.14 \\
0.01 & 0.08 & 0.76 \\
0.02 & 1.16 & 0.24
\end{pmatrix}, \quad \Psi = \begin{pmatrix}
7.79 & 0.61 & 0.40 \\
0.61 & 1.39 & 0.37 \\
0.40 & 0.37 & 0.64
\end{pmatrix}.
\]

(5.11)

The unique variance is: \( \text{diag}(U) = (0.34 \ 0.13 \ 0.16 \ 0.19 \ 0.29 \ 0.15 \ 0.21 \ 0.11 \ 0.21 \ 0.25 \ 0.15 \ 0.23 \ 0.00 \ 0.11 \ 0.19 \ 0.11) \). The percentage of explained common variance
is: 74.76. The explained common variances for all item-condition combinations are given in column 4 of Table 5.6. They are between 62 and 89 percent.

For interpreting (5.10) we compute $(C \otimes B)\Psi$; see columns 4, 5, 6 of Table 5.10. The solution (5.11) can be interpreted as follows. Factor 1 represents a general factor. Factor 2 represents item 5. Factor 3 represents mostly item 1. There is not much different between two conditions.

In the CP solution with $R = PQ = 3$ factors (see Table (4.4)), factor 1 is a combination of items 2,3,7,8 and has higher loadings for Yourself; factor 2 is a combination of items 1,4,5,6 and has higher loadings for Others; factor 3 is a contrast factor between Yourself and Others. This implies that solution (5.11) is not clearer than the CP solution. Additionally, solution (5.11) has less percent explained common variance than of the CP solution (74.76% versus 84.54%).

Clearly, to estimate the BJW data set, the three-mode CP factor model (4.9) is more useful than the three-mode Tucker3 factor model (5.2). Whereas, for the data set from Dickinson and Tice (1973), the three-mode Tucker3 factor model (5.2) is more useful. Hence, it depends on the data set, which model has a clearer interpretation.
Table 5.5: Values of \((C \otimes B) \Psi\) for the estimated Tucker3 covariance model (5.2) with \(P = 1\), \(Q \in \{2, 3\}\) factors fitted to the BJW data from section 4.3.2.

<table>
<thead>
<tr>
<th>variable</th>
<th>(P = 2, Q = 1)</th>
<th>(P = 3, Q = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>item 1, yourself</td>
<td>1.88 0.45</td>
<td>1.90 0.15 0.37</td>
</tr>
<tr>
<td>item 2, yourself</td>
<td>2.06 0.16</td>
<td>2.05 -0.05 0.08</td>
</tr>
<tr>
<td>item 3, yourself</td>
<td>1.94 0.05</td>
<td>1.94 -0.09 -0.09</td>
</tr>
<tr>
<td>item 4, yourself</td>
<td>2.00 0.64</td>
<td>2.04 0.38 0.25</td>
</tr>
<tr>
<td>item 5, yourself</td>
<td>1.26 1.04</td>
<td>1.34 0.89 0.23</td>
</tr>
<tr>
<td>item 6, yourself</td>
<td>2.06 0.68</td>
<td>2.10 0.42 0.24</td>
</tr>
<tr>
<td>item 7, yourself</td>
<td>1.84 0.03</td>
<td>1.83 -0.09 -0.09</td>
</tr>
<tr>
<td>item 8, yourself</td>
<td>2.01 0.18</td>
<td>2.02 0.02 -0.06</td>
</tr>
</tbody>
</table>

| item 1, others  | 1.94 0.47 | 1.98 0.16 0.38 |
| item 2, others  | 2.12 0.17 | 2.13 -0.06 0.09 |
| item 3, others  | 2.01 0.05 | 2.01 -0.09 -0.09 |
| item 4, others  | 2.06 0.66 | 2.13 0.40 0.26 |
| item 5, others  | 1.30 1.07 | 1.40 0.93 0.24 |
| item 6, others  | 2.13 0.70 | 2.19 0.44 0.25 |
| item 7, others  | 1.89 0.04 | 1.90 -0.10 -0.09 |
| item 8, others  | 2.08 0.19 | 2.10 0.03 -0.06 |

*Note:* Numbers larger than 0.5 are in boldfont.
Table 5.6: The explained common variances for all item-condition combinations and the communalities for the BJW data from section 4.3.2

<table>
<thead>
<tr>
<th>variable</th>
<th>(P = 2, Q = 1)</th>
<th>(P = 3, Q = 1)</th>
<th>(CP) with (R = 2)</th>
<th>(CP) with (R = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
<td>ECV% comm.</td>
</tr>
<tr>
<td>item 1, yourself</td>
<td>68 0.66</td>
<td>87 0.66</td>
<td>69 0.66</td>
<td>70 0.66</td>
</tr>
<tr>
<td>item 2, yourself</td>
<td>70 0.88</td>
<td>71 0.87</td>
<td>86 0.88</td>
<td>85 0.87</td>
</tr>
<tr>
<td>item 3, yourself</td>
<td>68 0.84</td>
<td>67 0.84</td>
<td>93 0.84</td>
<td>94 0.84</td>
</tr>
<tr>
<td>item 4, yourself</td>
<td>70 0.81</td>
<td>71 0.81</td>
<td>74 0.81</td>
<td>82 0.81</td>
</tr>
<tr>
<td>item 5, yourself</td>
<td>79 0.70</td>
<td>86 0.71</td>
<td>43 0.70</td>
<td>79 0.71</td>
</tr>
<tr>
<td>item 6, yourself</td>
<td>72 0.85</td>
<td>72 0.85</td>
<td>68 0.85</td>
<td>81 0.85</td>
</tr>
<tr>
<td>item 7, yourself</td>
<td>68 0.79</td>
<td>89 0.79</td>
<td>85 0.79</td>
<td>88 0.79</td>
</tr>
<tr>
<td>item 8, yourself</td>
<td>70 0.89</td>
<td>72 0.89</td>
<td>89 0.89</td>
<td>89 0.89</td>
</tr>
<tr>
<td>item 1, others</td>
<td>68 0.79</td>
<td>85 0.79</td>
<td>74 0.79</td>
<td>72 0.79</td>
</tr>
<tr>
<td>item 2, others</td>
<td>76 0.76</td>
<td>78 0.75</td>
<td>76 0.76</td>
<td>83 0.75</td>
</tr>
<tr>
<td>item 3, others</td>
<td>66 0.85</td>
<td>70 0.85</td>
<td>65 0.85</td>
<td>88 0.85</td>
</tr>
<tr>
<td>item 4, others</td>
<td>85 0.77</td>
<td>85 0.77</td>
<td>89 0.77</td>
<td>90 0.77</td>
</tr>
<tr>
<td>item 5, others</td>
<td>81 1.00</td>
<td>85 1.00</td>
<td>62 1.00</td>
<td>85 1.00</td>
</tr>
<tr>
<td>item 6, others</td>
<td>78 0.89</td>
<td>79 0.89</td>
<td>86 0.89</td>
<td>86 0.89</td>
</tr>
<tr>
<td>item 7, others</td>
<td>58 0.80</td>
<td>62 0.81</td>
<td>62 0.81</td>
<td>90 0.81</td>
</tr>
<tr>
<td>item 8, others</td>
<td>58 0.88</td>
<td>62 0.89</td>
<td>66 0.88</td>
<td>86 0.89</td>
</tr>
<tr>
<td>ECV% total</td>
<td>71.00</td>
<td>74.76</td>
<td>74.51</td>
<td>84.54</td>
</tr>
</tbody>
</table>

5.5 Discussion

In this chapter, we have proposed and demonstrated a method for three-mode factor analysis using MRFA to estimate unique variances \(U\) and Tucker3 to estimate the covariance matrix of the common part. This method is an extension of the CP-based method that is proposed in chapter 4.

By using the Tucker3 covariance model, the component matrices and the core array of interactions of the factors in the second and the third modes are not unique. Therefore we use the Joint Orthomax rotation of \(Kiers\ 1998a\) to rotate the Tucker3 solution to simple structure in matrices corresponding to the second and the third modes, and also in the core array.
The solutions in section 5.3 show that the Tucker3 solutions have a clearer interpretation than the CP solutions for the MTMM data of Dickinson and Tice (1973). However, the solutions in section 5.4 show that the CP solutions have a clearer interpretation than the Tucker3 solutions for the BJW data of section 4.3.2. Hence, the Tucker3 factor model is not always more appropriate than the CP factor model and vice versa. Therefore, we advise to use both models on the same dataset and to choose the model with the clearest interpretation (provided the explained common variances are similar in size).