Learning Structures and Time Series Homework 3

- Akankshi Mody, am92786

Question 1

a) $x_1 = 0.8\xi_1 + 0.4\xi_2 + \varepsilon_1$

$x_2 = 0.6\xi_1 + 0.6\xi_2 + \varepsilon_2$

$x_3 = 0.4\xi_1 + 0.8\xi_2 + \varepsilon_3$

loading matrix :>

$$\begin{array}{c|cc} & \xi_1 & \xi_2 \\ x_1 & \lambda_{11} & \lambda_{12} \\ x_2 & \lambda_{21} & \lambda_{22} \\ x_3 & \lambda_{31} & \lambda_{32} \end{array} = \begin{array}{c|cc} & \xi_1 & \xi_2 \\ x_1 & 0.8 & 0.4 \\ x_2 & 0.6 & 0.6 \\ x_3 & 0.4 & 0.8 \end{array}$$

we know, $\text{corr}(Y,X) = \dfrac{\text{cov}(Y,X)}{\sigma_Y \sigma_X}$

$\text{corr}(x_1, x_2) = \text{cov}(0.8\xi_1 + 0.4\xi_2 + \varepsilon_1,\ 0.6\xi_1 + 0.6\xi_2 + \varepsilon_2)$

$= 0.8 \times 0.6\ \text{cov}(\xi_1, \xi_2) + 0.8 \times 0.6\ \text{cov}(\xi_1, \xi_1)$
$+ 0.4 \times 0.6\ \text{cov}(\xi_2, \xi_1) + 0.4 \times 0.6\ \text{cov}(\xi_2, \xi_2)$
$+ 0.8\ \text{cov}(\xi_1, \varepsilon_2) + 0.4\ \text{cov}(\xi_2, \varepsilon_2)$
$+ 0.6\ \text{cov}(\varepsilon_1, \xi_1) + 0.6\ \text{cov}(\varepsilon_1, \xi_2) +$
$\text{cov}(\varepsilon_1, \varepsilon_2)$

$= 0.48 + 0.24 = 0.72$

Similarly, $\text{corr}(x_2, x_3) = 0.24 + 0.48 = 0.72$

$\text{corr}(x_3, x_1) = 0.32 + 0.32 = 0.64.$

Estimated loadings matrix :

	x_1	x_2	x_3
x_1	1	0.72	0.64
x_2	0.72	1	0.72
x_3	0.64	0.72	1

b) communality of $x_1 = (0.8)^2 + (0.4)^2 = 0.8$

$x_2 = (0.6)^2 + (0.6)^2 = 0.72$

$x_3 = (0.4)^2 + (0.8)^2 = 0.8$

Question 2

a) $Var(x_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \dots \lambda_{ip}^2 + Var(\epsilon_i) = 1$

This matrix from (point wise) square of FA loadings matrix gives $var(x_i)$ explained by ξ_i and λ_i

	ξ_1	ξ_2	Total var explained
x_1	0.64	0.16	0.8
x_2	0.36	0.36	0.72
x_3	0.16	0.64	0.8
Total var	1.16	1.16	2.32

b) No, it cannot be equivalent computationally to PCA - Unlike PCA, where number of PCs = number of variables and 100% variance is explained by all PCs, FA here has only 2 factors for 3 variables and only 2.32 out of 3 variance is explained.

Question 3

$$M = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix} \quad F = \text{factor pattern matrix}$$

a) $$F = \begin{bmatrix} 0.8 & 0.4 \\ 0.6 & 0.6 \\ 0.4 & 0.8 \end{bmatrix} \quad M = \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$$

$$FM = F^* = \begin{bmatrix} 0.85 & 0.28 \\ 0.85 & 0 \\ 0.85 & -0.28 \end{bmatrix}$$

$$F \times F^T = \begin{bmatrix} 0.8 & 0.4 \\ 0.6 & 0.6 \\ 0.4 & 0.8 \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 & 0.4 \\ 0.4 & 0.6 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8 & 0.72 & 0.64 \\ 0.72 & 0.72 & 0.72 \\ 0.64 & 0.72 & 0.8 \end{bmatrix}$$

$$FM \times M^T \times F^T = \begin{bmatrix} 0.85 & 0.28 \\ 0.85 & 0 \\ 0.85 & -0.28 \end{bmatrix} \begin{bmatrix} 0.7071 & 0.7071 \\ 0.7071 & -0.7071 \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 & 0.4 \\ 0.4 & 0.6 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8 & 0.72 & 0.64 \\ 0.72 & 0.72 & 0.72 \\ 0.64 & 0.72 & 0.8 \end{bmatrix}$$

Thus $F \times F^T = F \times M \times M^T \times F^T$

Also, $F \times M \times M^T \times F^T$ has corr matrix of

$$\begin{bmatrix} 1 & 0.72 & 0.64 \\ 0.72 & 1 & 0.72 \\ 0.64 & 0.72 & 1 \end{bmatrix}$$ (same as (a))

b) From F^*,

communality of x_1 $= (0.85)^2 + (0.28)^2 = 0.8$

$x_2 = (0.85)^2 + 0 = 0.72$

$x_3 = (0.85)^2 + (-0.28)^2 = 0.8$

F^* is orthonormal rotation of F, communality is same. Information of factors is still same after the rotation

Question 4

$X_1 = a\xi + \varepsilon_1$

$X_2 = b\xi + \varepsilon_2$

$X_3 = c\xi + \varepsilon_3$

a) For correlation to be same

$corr(X_1, X_2) = 0.72 = ab\ corr(\xi, \xi) = ab$ — (1)

$corr(X_2, X_3) = 0.72 = bc \cdot corr(\xi, \xi) = bc$ — (2)

$corr(X_3, X_1) = 0.64 = ac \cdot corr(\xi, \xi) = ac$ — (3)

From above equations, $b = \frac{0.72}{a} = \frac{0.72}{c} \Rightarrow a = c$

$ac = 0.64 \quad \therefore \quad a = c = 0.8$

$b = \frac{0.72}{0.8} = 0.9$

Hence, we see that its' possible for one-factor model to produce same corr. matrix as two factor model

b) 3(a) and 4(a), we can see that for a combination of manifest variables, there can be many sets of factors. and thus many sets of correlation matrices.

5. Run a "principal components" style of factor analysis to extract 6 factors.

a) Submit your SAS code.

```
PROC IMPORT DATAFILE="/home/u45121022/MSBA/EvaluateSupervisors.xlsx"
            OUT=WORK.EvaluateSupervisors
            DBMS=XLSX
            REPLACE;
RUN;

PROC FACTOR DATA = WORK.EvaluateSupervisors METHOD = principal PRIORS=ONE NFACTORS=6
MINEIGEN=0;
    VAR BEEFS PRIVILEGE NEWLEARN RAISES CRITICAL ADVANCE;
    TITLE "Principal Components Style Factor Analysis";
RUN;
```

Principal Components Style Factor Analysis

The FACTOR Procedure
Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 6 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	3.16922321	2.16287646	0.5282	0.5282
2	1.00634675	0.24343802	0.1677	0.6959
3	0.76290873	0.21039227	0.1272	0.8231
4	0.55251646	0.23526997	0.0921	0.9152
5	0.31724648	0.12548811	0.0529	0.9680
6	0.19175838		0.0320	1.0000

6 factors will be retained by the NFACTOR criterion.

Factor Pattern							
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
BEEFS	BEEFS	0.78219	-0.31363	0.38883	-0.23490	-0.10787	0.26797
PRIVILEGE	PRIVILEGE	0.70268	-0.30973	0.18990	0.60569	-0.02123	-0.08333
NEWLEARN	NEWLEARN	0.82140	-0.21777	-0.23756	-0.16709	0.43688	-0.05153
RAISES	RAISES	0.87704	0.11590	0.00490	-0.27139	-0.25930	-0.27649
CRITICAL	CRITICAL	0.40022	0.80479	0.39938	0.07429	0.16271	0.02533
ADVANCE	ADVANCE	0.67791	0.32172	-0.59975	0.15293	-0.14347	0.18237

b) How many factors would you retain? Why?

The eigenvalue for the first two factors is more than 1. By the Kaiser rule, I would retain 2 factors.

c) Try to interpret the first two factors.

Factor 1 exhibits high weightage from all the variables except for "Critical". This means that an employee who has a high score on Factor 1 likely has high scores on all other variables in the dataset – i.e. on handling complaints well, special privileges, opportunity to learn new things, raises based on merit, and advances employees to better jobs. But this employee is not critical. Such an employee would be a valuable asset to a company and would be liked well by employees working under him/her.

Factor 2 has a high score only for "Critical". Such a supervisor will be considered critical and will probably be disliked by employees working under him/her.

6. How well do the six factors of question 5 explain the OVERALL supervisor rating – individually and collectively?

```
PROC REG DATA=WORK.prinfactors;
    model OVERALL = Factor1 Factor2 Factor3 Factor4 Factor5 Factor6 /ss1 ss2;
RUN;
```

The REG Procedure
Model: MODEL1
Dependent Variable: OVERALL OVERALL

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	3147.96634	524.66106	10.50	<.0001
Error	23	1149.00032	49.95654		
Corrected Total	29	4296.96667			

Root MSE	7.06799	R-Square	0.7326
Dependent Mean	64.63333	Adj R-Sq	0.6628
Coeff Var	10.93552		

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS
Intercept	Intercept	1	64.63333	1.29043	50.09	<.0001	125324	125324
Factor1		1	8.22970	1.31249	6.27	<.0001	1964.11131	1964.11131
Factor2		1	-3.41693	1.31249	-2.60	0.0159	338.58649	338.58649
Factor3		1	3.60689	1.31249	2.75	0.0115	377.27898	377.27898
Factor4		1	-3.63138	1.31249	-2.77	0.0110	382.41999	382.41999
Factor5		1	0.94266	1.31249	0.72	0.4799	25.76965	25.76965
Factor6		1	1.43599	1.31249	1.09	0.2852	59.79991	59.79991

Collective: We run a regression with all the factors to determine the overall rating. The six factors then explain 73.26% of the OVERALL rating. From the output we see that Factors 5 and 6 are insignificant.

Individual: Variance explained is be given by R-square * Type I SS / SSE.

Factors	Variance explained (%)
Factor 1	45.70912743
Factor 2	7.879641514
Factor 3	8.780099623

Factor 4	8.899742069
Factor 5	0.599715612
Factor 6	1.391673523
Total	73.25999977

7. Run a principal factor analysis, with R-square type initial estimates of communalities, to extract six factors.

a) Submit your SAS code.

```
PROC FACTOR DATA = WORK.EvaluateSupervisors METHOD = principal PRIORS=SMC NFACTORS=6
MINEIGEN=0
OUT=WORK.prinfactors;
    VAR BEEFS PRIVILEGE NEWLEARN RAISES CRITICAL ADVANCE;
RUN;
```

The FACTOR Procedure
Initial Factor Method: Principal Factors

Prior Communality Estimates: SMC					
BEEFS	PRIVILEGE	NEWLEARN	RAISES	CRITICAL	ADVANCE
0.62505535	0.37534801	0.55967371	0.67513759	0.18574001	0.48759750

Eigenvalues of the Reduced Correlation Matrix: Total = 2.90855217 Average = 0.48475869				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.71771612	2.31678131	0.9344	0.9344
2	0.40093481	0.22202808	0.1378	1.0722
3	0.17890672	0.18294128	0.0615	1.1337
4	-.00403456	0.16150739	-0.0014	1.1324
5	-.16554195	0.05388702	-0.0569	1.0754
6	-.21942897		-0.0754	1.0000

3 factors will be retained by the MINEIGEN criterion.

Factor Pattern				
		Factor1	Factor2	Factor3
BEEFS	BEEFS	0.74755	-0.36273	0.12483
PRIVILEGE	PRIVILEGE	0.61091	-0.17725	-0.06404
NEWLEARN	NEWLEARN	0.76629	-0.05146	-0.21483
RAISES	RAISES	0.84947	0.11042	0.13720
CRITICAL	CRITICAL	0.32091	0.25308	0.26760
ADVANCE	ADVANCE	0.61147	0.39882	-0.15045

Only the first 3 factors will be retained by since the eigenvalues for the last 3 factors are negative.

b) Try to interpret the first two factors.

Factor 1 exhibits high weightage from all the variables except for “Critical”. This means that an employee who has a high score on Factor 1 likely has high scores on all other variables in the dataset –

i.e. on handling complaints well, special privileges, opportunity to learn new things, raises based on merit, and advances employees to better jobs. But this employee is not critical. Such an employee would be a valuable asset to a company and would be liked well by employees working under him/her.

Factor 2 exhibits low weightage on all variables. Also, it has negative loadings on the first 3 variables. Such a supervisor probably does not handle employee complaints well, does not give them privilege and opportunity to learn. He/she is critical and has little ability to advance employee to better jobs.

c) Succinctly compare the factor analysis of Q7 with the factor analysis of Q5.

In Q7, the total number of retained factors is 3 whereas in Q5, all the factors had with positive eigenvalues. The variance being explained by has also reduced.

8. Run a principal factor analysis, with R-square type initial estimates of communalities, to extract two factors, apply a varimax rotation to the initial factor pattern, and add factor scores for all 30 supervisors to a dataset called WORK.EvaluateSupervisors_scores (including all original data as well).
a) Submit your SAS code.

```
PROC FACTOR DATA = WORK.EvaluateSupervisors METHOD = principal PRIORS=SMC NFACTORS=2
ROTATE=varimax
OUT=WORK.EvaluateSupervisors_scores;
    VAR BEEFS PRIVILEGE NEWLEARN RAISES CRITICAL ADVANCE;
RUN;
```

The FACTOR Procedure
Initial Factor Method: Principal Factors

Prior Communality Estimates: SMC					
BEEFS	PRIVILEGE	NEWLEARN	RAISES	CRITICAL	ADVANCE
0.62505535	0.37534801	0.55967371	0.67513759	0.18574001	0.48759750

Eigenvalues of the Reduced Correlation Matrix: Total = 2.90855217 Average = 0.48475869				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.71771612	2.31678131	0.9344	0.9344
2	0.40093481	0.22202808	0.1378	1.0722
3	0.17890672	0.18294128	0.0615	1.1337
4	-.00403456	0.16150739	-0.0014	1.1324
5	-.16554195	0.05388702	-0.0569	1.0754
6	-.21942897		-0.0754	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern			
		Factor1	Factor2
BEEFS	BEEFS	0.74755	-0.36273
PRIVILEGE	PRIVILEGE	0.61091	-0.17725
NEWLEARN	NEWLEARN	0.76629	-0.05146
RAISES	RAISES	0.84947	0.11042
CRITICAL	CRITICAL	0.32091	0.25308
ADVANCE	ADVANCE	0.61147	0.39882

b) Are the first two initial factors the same as the first two factors in the preceding question?

Yes, they are the same.

c) Verify that the varimax factor rotation matrix is orthonormal.

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix		
	1	2
1	0.79912	0.60117
2	-0.60117	0.79912

The magnitude of each vector:

(0.79912)^2 + (0.60117)^2 = 0.638592774 + 0.361405369 = 1

(0.60117)^2 + (0.79912)^2 = 0.361405369 + 0.638592774 = 1

Thus, the vectors are normalized.

Dot Product:

0.79912 * 0.60117 + 0.79912* -0.60117 = 0

Thus, the vectors are orthogonal.

Thus, the varimax factor rotation matrix is orthonormal.

d) Try to interpret the varimax rotated factor pattern.

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix		
	1	2
1	0.79912	0.60117
2	-0.60117	0.79912

Rotated Factor Pattern			
		Factor1	Factor2
BEEFS	BEEFS	0.81544	0.15954
PRIVILEGE	PRIVILEGE	0.59475	0.22561
NEWLEARN	NEWLEARN	0.64329	0.41955
RAISES	RAISES	0.61245	0.59891
CRITICAL	CRITICAL	0.10431	0.39516
ADVANCE	ADVANCE	0.24889	0.68630

Variance Explained by Each Factor	
Factor1	Factor2
1.8804192	1.2382317

Final Communality Estimates: Total = 3.118651					
BEEFS	PRIVILEGE	NEWLEARN	RAISES	CRITICAL	ADVANCE
0.69039962	0.40462377	0.58984535	0.73379378	0.16703250	0.53295590

Factor exhibits high loading on BEEFS, PRIVILEGE, NEWLEARN, RAISES and low loadings on CRITICAL AND ADVANCE. Such a supervisor is not critical and does not have the ability to advance employee to better jobs.

Factor 2 exhibits low but positive loadings on all variables with the highest loadings are on ADVANCE AND RAISES. Such a supervisor would have the ability to advance employee to better jobs as well as giving raises.

9. Calculate means, standard deviations, and correlation for the factor scores of the 30 supervisors of the varimax rotated factors that you stored in WORK.EvaluateSupervisors_scores in the preceding question. [You may use PROC CORR, which includes PROC MEANS output by default.]

a) Given the assumptions of factor analysis, do your summary statistics surprise you? Why or why not?

```
PROC CORR DATA=WORK.EvaluateSupervisors_scores;
RUN;
```

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
OVERALL	30	64.63333	12.17256	1939	40.00000	85.00000	OVERALL
BEEFS	30	66.60000	13.31476	1998	37.00000	90.00000	BEEFS
PRIVILEGE	30	53.13333	12.23543	1594	30.00000	83.00000	PRIVILEGE
NEWLEARN	30	56.36667	11.73701	1691	34.00000	75.00000	NEWLEARN
RAISES	30	64.63333	10.39723	1939	43.00000	88.00000	RAISES
CRITICAL	30	74.76667	9.89491	2243	49.00000	92.00000	CRITICAL
ADVANCE	30	42.93333	10.28871	1288	25.00000	72.00000	ADVANCE
Factor1	30	0	0.85285	0	-1.86118	1.36525	
Factor2	30	0	0.77645	0	-1.31002	2.01667	

Pearson Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0									
	OVERALL	BEEFS	PRIVILEGE	NEWLEARN	RAISES	CRITICAL	ADVANCE	Factor1	Factor2
OVERALL OVERALL	1.00000	0.82542 <.0001	0.42612 0.0189	0.62368 0.0002	0.59014 0.0006	0.15644 0.4091	0.15509 0.4132	0.82349 <.0001	0.19094 0.3122
BEEFS BEEFS	0.82542 <.0001	1.00000	0.55829 0.0013	0.59674 0.0005	0.66920 <.0001	0.18771 0.3205	0.22458 0.2328	0.95614 <.0001	0.20547 0.2760
PRIVILEGE PRIVILEGE	0.42612 0.0189	0.55829 0.0013	1.00000	0.49333 0.0056	0.44548 0.0136	0.14723 0.4375	0.34329 0.0633	0.69736 <.0001	0.29057 0.1193
NEWLEARN NEWLEARN	0.62368 0.0002	0.59674 0.0005	0.49333 0.0056	1.00000	0.64031 0.0001	0.11597 0.5417	0.53162 0.0025	0.75429 <.0001	0.54034 0.0021
RAISES RAISES	0.59014 0.0006	0.66920 <.0001	0.44548 0.0136	0.64031 0.0001	1.00000	0.37688 0.0401	0.57419 0.0009	0.71812 <.0001	0.77135 <.0001
CRITICAL CRITICAL	0.15644 0.4091	0.18771 0.3205	0.14723 0.4375	0.11597 0.5417	0.37688 0.0401	1.00000	0.28334 0.1292	0.12230 0.5197	0.50894 0.0041
ADVANCE ADVANCE	0.15509 0.4132	0.22458 0.2328	0.34329 0.0633	0.53162 0.0025	0.57419 0.0009	0.28334 0.1292	1.00000	0.29183 0.1176	0.88390 <.0001
Factor1	0.82349 <.0001	0.95614 <.0001	0.69736 <.0001	0.75429 <.0001	0.71812 <.0001	0.12230 0.5197	0.29183 0.1176	1.00000	0.29041 0.1195
Factor2	0.19094 0.3122	0.20547 0.2760	0.29057 0.1193	0.54034 0.0021	0.77135 <.0001	0.50894 0.0041	0.88390 <.0001	0.29041 0.1195	1.00000

The means of Factors 1 and2 do not surprise me since they match our assumptions. But, the standard deviations of Factors should be 1, since the data is standardized but they are not. Even the correlations of the factors are supposed to be different in the basic Factor Analysis model.

10. Extract the default number of factors by the maximum likelihood method with R-square type initial estimates of communalities. You may need to use the ULTRAHEYWOOD or HEYWOOD option to deal with estimated communalities that exceed 1. You may assume that the assumptions of the maximum likelihood model apply.

a) Does the output provide support for the hypothesis that common factors exist?

```
PROC FACTOR DATA = WORK.Evaluatesupervisors METHOD=ML PRIORS=SMC ULTRAHEYWOOD;
    VAR BEEFS PRIVILEGE NEWLEARN RAISES CRITICAL ADVANCE;
RUN;
```

Significance Tests Based on 30 Observations			
Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	15	65.5127	<.0001
HA: At least one common factor			
H0: 2 Factors are sufficient	4	2.8155	0.5892
HA: More factors are needed			

We can see that the p-value for no common factors is less than 0.05. Hence, we can reject this null hypothesis indicating that the data is has at least one common factor.

b) Does the output provide support for the hypothesis that the default number of extracted factors is adequate?

We can see that the p-value is 0.5892. Hence, we fail to reject the null hypothesis can extract the sufficient number of factors which is 2.