STA2101f19 Assignment Nine*

Ahmed Nawaz Amanullah †

April 2021

1

1.1 Problem

A farm co-operative (co-op) is an association of farmers. The co-op can buy fertilizer and other supplies in large quantities for a lower price, it often provides a common storage location for harvested crops, and it arranges sale of farm products in large quantities to grocery chains and other food supplies. Farm co-ops usually have professional managers, and some do a better job than others.
We have data from a study of farm co-op managers. The variables in the "latent variable" part of the model are the following, but note that one of them is assumed observable.

- Knowledge of business principles and products (economics, fertilizers and chemicals). This is a latent variable measured by **know1** and **know2**.

- Profit-loss orientation ("Tendency to rationally evaluate means to an economic end"). This is a latent variable measured by **ploss1** and **ploss2**.

†with help from the Overleaf team

- Job satisfaction. This is a latent variable measured by **sat1** and **sat2**.
- Formal Education(.?) This is an observable variable, assumed to be measured without error.
- Job performance. This is a latent variable measured by **perf1** and **perf2**.

The data file has these observable variables in addition to an identification code for the managers.

know1: Knowledge measurement 1

know2: Knowledge measurement 2

ploss1: Profit-Loss Orientation 1

ploss2: Profit-Loss Orientation 2

sat1: Job Satisfaction 1

sat2: Job Satisfaction 2

educat: Number of years of formal schooling divided by 6.

perf1: Job Performance 1

perf2: Job Performance 2

In this study, the double measurements are obtained by just splitting questionnaires in two, as in split half reliability. Furthermore, all the measurement errors are assumed independent of one another. This is consistent with mainstream psychometric theory, though maybe not with common sense. For this assignment, please assume that the errors are independent of one another, and independent of the exogenous variables. The explanatory variables, of course, should *not* be assumed independent of one another.

In the two main published analyses of these data, the latent exogenous variables were knowledge, profit-loss orientation, education and job satisfaction. The latent response variable was job performance. However, let's make it more interesting. Let's say that the latent exogenous variables are knowledge, education and profit-loss orientation, and that these influence job performance (possibly with a zero regression coefficient; we can test that). Job performance is also influenced by job satisfaction. Job satisfaction, in turn, is influenced by job performance (it feels good to do a good job), but not directly by any of the exogenous variables. So job satisfaction is endogenous too.

(a) Please make a path diagram. put (Put?) Greek letters on all the arrows, including curved arrows, unless the coefficient is one.

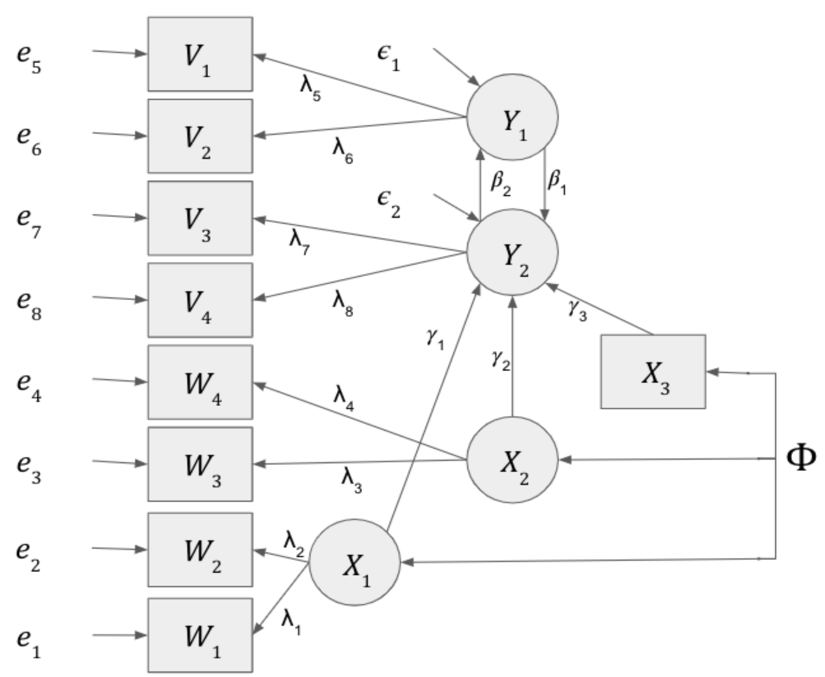
(b) List the parameters that appear in the covariance matrix of the observable data.

(c) Does this model pass the test of the parameter count rule? Answer Yes or no and give the numbers.

The parameters of this model are identifiable in most of the parameter space. Details will be taken up in class.

1.2 Solution

(a) Path diagram disclosed below



In this path diagram,

- X_1: Knowledge of business principles and products (economics, fertilizers and chemicals). This is a latent variable measured by **know1** and **know2** (i.e. W_1 and W_2 respectively).
- X_2: Profit-loss orientation ("Tendency to rationally evaluate means to an economic end"). This is a latent variable measured by **ploss1** and **ploss2** (i.e. W_3 and W_4 respectively).
- Y_1: Job satisfaction. This is a latent variable measured by **sat1** and **sat2** (i.e. V_1 and V_2 respectively).
- X_3: Formal Education. This is an observable variable, assumed to be measured without error.
- Y_2: Job performance. This is a latent variable measured by **perf1** and **perf2** (i.e. V_3 and V_4 respectively).

(b) Let's go about this systematically, from left to right of the path diagram,

- Measurement error variances, $Var(e_i) = \omega_i$ for $i = 1, 2, ..., 8$. This gives us 8 parameters.
- Factor loadings, λ_i for $i = 1, 2, ..., 8$. This gives us 8 parameters.
- Regression error variances, $Var(\epsilon_i) = \psi_i$ for $i = 1, 2$. This gives us 2 parameters.
- Regression coefficients, γ_i for $i = 1, 2, 3$ and β_i for $i = 1, 2$. This gives us 5 parameters.
- Explanatory variables unique variances and covariances. Let

$$cov \begin{pmatrix} X_{i,1} \\ X_{i,2} \\ X_{i,3} \end{pmatrix} = \left(\begin{array}{c|c|c} \phi_{11} & \phi_{12} & \phi_{13} \\ \hline & \phi_{22} & \phi_{23} \\ \hline & & \phi_{33} \end{array}\right) = \mathbf{\Phi}. \tag{1}$$

  This gives us $\frac{3(3+1)}{2} = 6$ parameters.

So, in total we have 29 parameters.

(c) In our response to **1 (b)** we have established that there are 29 parameters appearing in the observable data covariance matrix. The observable data covariance matrix gives us $\frac{9(9+1)}{2} = 45$ covariance structure equations (i.e. the number of unique variances and covariances in it). Since number of covariance structure equations exceeds number of parameters appearing in the observable data covariance matrix, **Yes**, this model passes the test of the parameter count rule.

2

2.1 Problem

The file co-opManager.data.txt has raw data for the study described in Question 1. This is a reconstructed data set based on a covariance matrix in Jorekog (1978, p.465). Joreskog got it from Warren, White and Fuller (1974).

Using **lavaan**, fit the model in your path diagram and look at **summary**. There are 98 co-ops, so please make sure you are reading the correct number of cases. For comparison, my value of the G^2 test statistic for model fit is 29.357. If you got this, we must be fitting the same model.

(a) Based on the number of covariance structure equations and the number of unknown parameters, how many equality restrictions should the model impose on the covariance matrix? The answer is a single number; fortunately, you need not say exactly what the equality restrictions are.

(b) Does your model fit the data adequately? Answer Yes or No and give three numbers: a chisquared statistic, the degrees of freedom, and a p-value. The degrees of freedom should agree with your answer to Question 2(a)i (2(a)?).

(c) In plain, non-statistical language, what are the main conclusions of this study? Be able to back up your conclusions with hypothesis tests that reject H_0 at $\alpha = 0.05$. Of course you keep quiet about it in your plain-language conclusions.

(d) It's remarkable that one can assess the effect of satisfaction on performance *and* the effect of performance on satisfaction. Be able to give the value of the test statistics and the p-values. It's a little disappointing, but these data are a re-creation of a real data set. Measurable job satisfaction is notoriously unrelated to any actual behavior - unless that behaviour consists of more talk.

(e) Carry out a Wald test of all the regression coefficients in the latent variable model at once; I count three γ_j and two β_j. Be able to give the value of the chi-squared test statistic, the degrees of freedom, and the p-value - all numbers from your printout. Using the usual $\alpha = 0.05$ significance level, is there evidence that at lease one regression coefficient must be non-zero? You can tell which ones from the output of **summary**.

(f) Estimate the reliability of knowledge measure one and knowledge measure two; give 95% confidence intervals as well. There is an easy way to do this. I almost asked about the reliability of job satisfaction, which is a nightmare for this model.

(g) Test whether the reliabilities of the two knowledge measures are equal; as you know, this is equivalent to testing quality of the measurement error variances. Be able to give the value of the test statistic, the p-value, and a draw a directional conclusion if one is warranted.

(h) There is another way to estimate reliability. Suppose that $D_1 = F + e_1$ and $D_2 = F + e_2$. If $Var(e_1) = Var(e_2) = \omega$, we call the measurements "equivalent," ("equivalent",?) and their common reliability is $\phi/(\omega + \phi)$. Calculate $Corr(D_1, D_2)$. This suggests a sample correlation as an estimate of reliability.

(i) Use the **cor** function to get a sample correlation matrix of all the observable variables. Assuming the measurements of knowledge are equivalent, can you find another estimate of the common reliability? How does it compare to your earlier estimates?

2.2 Solution

The R code for this question is disclosed below,

```
# Author: Ahmed N Amanullah
# St Num: 1001325773
# STA2101 Assignment 9 Ques 2

# first read the data
# first read data
```

```r
d_source <- "http://www.utstat.toronto.edu/~brunner/data/legal/"
d_source <- paste(d_source, "co-opManager.data.txt",sep="")
mdata <- read.table(d_source, header = T)

# see if data loaded alright
head(mdata)
summary(mdata)
dim(mdata)

# dim returns: 98  9
# so all cases are being read

# ok, now bigger fish to fry...
# try to fit model with lavaan
# load library
library(lavaan)

# model string
mmodel =
        ##################################################
    '
        # Latent variable model
        # ------------------------
        perf ~ gamma1*know + gamma2*ploss + gamma3*educat + beta21*sat
        sat ~ beta12*perf
        #
        # Measurement model
        # --------------------
        know =~ lambda1*know1 + lambda2*know2
        ploss =~ lambda3*ploss1 + lambda4*ploss2
        perf =~ lambda5*perf1 + lambda6*perf2
        sat =~ lambda7*sat1 + lambda8*sat2
        #
        # Variances and Covariances
        # ------------------------------
        # Of latent explanatory variables
        know ~~ phix11*know; know ~~ phix12*ploss; know ~~ phix13*educat
                                ploss ~~ phix22*ploss; ploss ~~ phix23*educat
```

```
                                        educat ~~ phix33*educat
        # of errors terms in latent regression
        perf ~~ psi1*perf
        sat ~~ psi2*sat
        # of measurement errors
        know1 ~~ omega1*know1
        know2 ~~ omega2*know2
        ploss1 ~~ omega3*ploss1
        ploss2 ~~ omega4*ploss2
        perf1 ~~ omega5*perf1
        perf2 ~~ omega6*perf2
        sat1 ~~ omega7*sat1
        sat2 ~~ omega8*sat2
        # Bounds (variances are postive)
        phix11 > 0; phix22 > 0; phix33 > 0
        psi1 > 0; psi2 > 0
        omega1 > 0; omega2 > 0; omega3 > 0; omega4 > 0
        omega5 > 0; omega6 > 0; omega7 > 0; omega8 > 0
        '############ End of mmodel ##############

# model string ready, now to see if it fits...
# Moment of truth... the palpitations....
# fit model
fit1 = lavaan(mmodel, data = mdata)

# display summary
summary(fit1)

# display log likelihood
logLik(fit1)

# Model does not fit... issue with standard error
# Try a model with no factor loadings...
# model string
mmodel2 =
        ################################################
    '
        # Latent variable model
```

```
        # ---------------------------------
        perf ~ gamma1*know + gamma2*ploss + gamma3*educat + beta12*sat
        sat ~ beta21*perf
        #
        # Measurement model
        # ---------------------------
        know =~ 1*know1 + 1*know2
        ploss =~ 1*ploss1 + 1*ploss2
        perf =~ 1*perf1 + 1*perf2
        sat =~ 1*sat1 + 1*sat2
        #
        # Variances and Covariances
        # ----------------------------------------
        # Of latent explanatory variables
        know ~~ phix11*know; know ~~ phix12*ploss; know ~~ phix13*educat
                                ploss ~~ phix22*ploss; ploss ~~ phix23*educat
                                                                        educat ~~ phix33*educat
        # of errors terms in latent regression
        perf ~~ psi1*perf
        sat ~~ psi2*sat
        # of measurement errors
        know1 ~~ omega1*know1
        know2 ~~ omega2*know2
        ploss1 ~~ omega3*ploss1
        ploss2 ~~ omega4*ploss2
        perf1 ~~ omega5*perf1
        perf2 ~~ omega6*perf2
        sat1 ~~ omega7*sat1
        sat2 ~~ omega8*sat2
        # Bounds (variances are postive)
        phix11 > 0; phix22 > 0; phix33 > 0
        psi1 > 0; psi2 > 0
        omega1 > 0; omega2 > 0; omega3 > 0; omega4 > 0
        omega5 > 0; omega6 > 0; omega7 > 0; omega8 > 0
        '########### End of mmodel ##############

# Second try
# fit model
```

```
fit2 = lavaan(mmodel2, data = mdata)

# display summary
summary(fit2)

# display log likelihood
logLik(fit2)

# Ok this one fits, and the G^2 matches the prof's (29.357),
#       so keep this. p-value is 0.207

# Try one with half the loadings just for funsies
# model string
mmodel3 =
  ###############################################
  '
 # Latent variable model
 # ----------------------
 perf ~ gamma1*know + gamma2*ploss + gamma3*educat + beta21*sat
 sat ~ beta12*perf
 #
 # Measurement model
 # ------------------
 know =~ 1*know1 + lambda2*know2
 ploss =~ 1*ploss1 + lambda4*ploss2
 perf =~ 1*perf1 + lambda6*perf2
 sat =~ 1*sat1 + lambda8*sat2
 #
 # Variances and Covariances
 # --------------------------
 # Of latent explanatory variables
 know ~~ phix11*know; know ~~ phix12*ploss; know ~~ phix13*educat
         ploss ~~ phix22*ploss; ploss ~~ phix23*educat
                         educat ~~ phix33*educat
 # of errors terms in latent regression
 perf ~~ psi1*perf
 sat ~~ psi2*sat
 # of measurement errors
```

```
 know1 ~~ omega1*know1
 know2 ~~ omega2*know2
 ploss1 ~~ omega3*ploss1
 ploss2 ~~ omega4*ploss2
 perf1 ~~ omega5*perf1
 perf2 ~~ omega6*perf2
 sat1 ~~ omega7*sat1
 sat2 ~~ omega8*sat2
 # Bounds (variances are postive)
 phix11 > 0; phix22 > 0; phix33 > 0
 psi1 > 0; psi2 > 0
 omega1 > 0; omega2 > 0; omega3 > 0; omega4 > 0
 omega5 > 0; omega6 > 0; omega7 > 0; omega8 > 0
 '############ End of mmodel ##############

# Second try
# fit model
fit3 = lavaan(mmodel3, data = mdata)

# display summary
summary(fit3)

# display log likelihood
logLik(fit3)

# Interestingly this one seems to fit too but with different G^2
#       than prof's (21.505); p-value is higher too (0.368)

# will continue analysis of question with fit2 as the G^2 for that
#       matches with prof's, so prof is porbably using that model...

# parts (a)-(d) answerable from output

# part (e) - Wald test for all regression coefficients
# first borrow Wtest function written by Prof. Brunner
source("http://www.utstat.toronto.edu/~brunner/Rfunctions/Wtest.txt")

# retrieve coefficients and asymptotic covariances of parameters
```

```
mcoef <- coef(fit2);mcoef
mvcov <- vcov(fit2);mvcov

# get number of parameters in model
nparam <- length(mcoef);nparam

# by inspecting mcoef, we see the regression coefficients occupy
#       the first 5 elements
# so L matrix will have 5 rows - one for each hypothesis
#       and the first 5 columns will be an identity matrix, with
#       the remaining 16 columns all 0
LL <- cbind(diag(5),matrix(rep(0, (nparam-5)*5),nrow=5)); LL

# now apply Wald test
Wtest(LL,   mcoef,  mvcov)

# part (f): The reliability estimates and 95% CI of the knowledge
#       measures - can use lavaan for this:
mmodel2b =
 ##################################################
 '
 # Latent variable model
 # -----------------------
 perf ~ gamma1*know + gamma2*ploss + gamma3*educat + beta12*sat
 sat ~ beta21*perf
 #
 # Measurement model
 # -------------------
 know =~ 1*know1 + 1*know2
 ploss =~ 1*ploss1 + 1*ploss2
 perf =~ 1*perf1 + 1*perf2
 sat =~ 1*sat1 + 1*sat2
 #
 # Variances and Covariances
 # ----------------------------
 # Of latent explanatory variables
 know ~~ phix11*know; know ~~ phix12*ploss; know ~~ phix13*educat
         ploss ~~ phix22*ploss; ploss ~~ phix23*educat
```

```
                         educat ~~ phix33*educat
 # of errors terms in latent regression
 perf ~~ psi1*perf
 sat ~~ psi2*sat
 # of measurement errors
 know1 ~~ omega1*know1
 know2 ~~ omega2*know2
 ploss1 ~~ omega3*ploss1
 ploss2 ~~ omega4*ploss2
 perf1 ~~ omega5*perf1
 perf2 ~~ omega6*perf2
 sat1 ~~ omega7*sat1
 sat2 ~~ omega8*sat2
 # Bounds (variances are postive)
 phix11 > 0; phix22 > 0; phix33 > 0
 psi1 > 0; psi2 > 0
 omega1 > 0; omega2 > 0; omega3 > 0; omega4 > 0
 omega5 > 0; omega6 > 0; omega7 > 0; omega8 > 0
 # Parameter functions
 relk1 := phix11/(phix11+omega1)
 relk2 := phix11/(phix11+omega2)
 '############ End of mmodel ##############

# fit model, display parameter estimates
fit2b = lavaan(mmodel2b, data = mdata)

# display summary
summary(fit2b)

# display log likelihood
logLik(fit2b)

# param ests
pEsts <- parameterEstimates(fit2b);pEsts

# Just display output for the rels
pEsts[30:31,c(1,5,9,10)]
# the above gives the reliabilities ests and their CIs
```

```
# Try out delta method too
# first the reliability est
relk1 <- mcoef[6]/(mcoef[6] + mcoef[14]); relk1
# ok good so far

# delta method
n <- dim(mdata)[1]; n # sample size
# gdot
gdot <- mcoef[14]/((mcoef[6] + mcoef[14])^2)
gdot <- c(gdot, -mcoef[6]/((mcoef[6] + mcoef[14])^2))
gdot <- matrix(gdot, nrow=1); gdot
# asymptotic covariance
Sigma <- c(mvcov[6,6], mvcov[6,14], mvcov[6,14], mvcov[14,14])
Sigma <- matrix(Sigma, nrow = 2); Sigma
# now get the asymptotic variance of relk1:
v_relk1 <- (1/n)*(gdot%*%Sigma%*%t(gdot)); v_relk1
se_relk1 <- sqrt(v_relk1); se_relk1

# delta method se doesn't match se from parameterEstimates
#       (difference of factor of 10!)
# for now, will use answers from lavaan as it is the simple way
#       as the question asks + I maybe doing something wrong...

# part (g): Test Ho: omega1 = omega2 (i.e. reliabilities of
#       knowledge measures being equal)
# Use Wtest()
# L matrix will have 1 and -1 in positions of omega1 and omega2
# These are columns 14 and 15 in the coef() output...
LL<- matrix(rep(0, 13),nrow=1)
LL <- cbind(LL,1,-1,matrix(rep(0, nparam-15),nrow=1)); LL
# apply Wtest:
Wtest(LL,  mcoef, mvcov)

# part (h) - Work done on paper; no R needed
# correlation compt assuming measurement not equal
cov_d1d2 <- mcoef[6]
vard1_vard2 <- (sqrt((mcoef[6]+mcoef[14])*(mcoef[6]+mcoef[15])))
```

```r
corr_d1d2 <- cov_d1d2/vard1_vard2; corr_d1d2


# part (i) - generate correlation matrix of observable data
mcor <- cor(mdata);  mcor
```

We executed this code in R console and extracted the responses for this question from the output observed. Note that in our execution we confirmed that the correct number of cases is being read. Furthermore, our model from question 1 did not fit. However, the model fitted when we excluded the factor loadings, and moreover, the G^2 test statistic for this model (i.e. excluding the factor models or setting them to one specifically) matched what the professor got. So we will use this model for the sub-questions of this question.

(a) From our responses to 1(b) and 1(c), we have 29 parameters appearing in the observable data covariance matrix and 45 covariance structure equations from the observable data covariance matrix. However, since model fit we are using is the one without the factor loadings (i.e. factor loadings set to 1), there are $29 - 8 = 21$ parameters appearing in the observable data covariance matrix. So the model imposes $45 - 21 = 24$ equality restrictions on the covariance matrix.

(b) $G^2 = 29.357$, $df = 24$ and p-value of 0.207. Since p-value is greater than 0.05, **Yes**, model fits the data adequately. Furthermore the degrees of freedom matches the number of equality restrictions we obtained in our response to the previous sub-question.

(c) We have evidence to suggest that knowledge and profit-loss orientation is positively related to job performance. However, there is insufficient evidence to suggest that education and job satisfaction (either direction for the latter) is related with job performance.

(d)
- β_1 (effect of satisfaction on performance): $Z = 1.117$ and p-value of 0.264.
- β_2 (effect of performance on satisfaction): $Z = -0.126$ and p-value of 0.900.

As mentioned in our response to the previous sub-question, there is insufficient evidence to suggest that job satisfaction (either direction) is related with job performance.

(e) Borrowing the professor's written **Wtest()** function for carrying out this Wald test. We are testing the regression coefficients, of which there are 5, being 0. So our **L** matrix will consist of 5 rows, one for each coefficient with the column corresponding to the coefficient set to 1 (and the other columns set to 0 for that row). Since we have 5 (non-redundant) equalities in our hypothesis, we have 5 degrees of freedom.

After carrying out the test, we have test statistic $W = 6.351135e + 01$, $df = 5$ and p-value of $2.281064e - 12$. Since p-value is less than the usual $\alpha = 0.05$ significance level, there is evidence to suggest that at least one regression coefficient is non-zero. As mentioned, we can tell which ones are non-zero from the **summary** output, specifically the Z-tests generated by it for the coefficients (the findings of which are disclosed in our response to 2(c)).

(f) Recall that reliability of a measurement is the proportion of its variance that does *not* come from the measurement error. In this question our focus is on the measures of the knowledge latent variable. The variance of these measures are

$$\begin{aligned}
Var(W_1) &= Var(X_1 + e_1) && (2)\\
&= Cov(X_1 + e_1, X_1 + e_1) && (3)\\
&= Cov(X_1 + e_1, X_1) + Cov(X_1 + e_1, e_1) && (4)\\
&= Cov(X_1 + e_1, X_1) + Cov(X_1, e_1) + Cov(e_1, e_1) && (5)\\
&= Cov(X_1 + e_1, X_1) + 0 + \omega_1 && (6)\\
&= Cov(X_1, X_1) + Cov(e_1, X_1) + \omega_1 && (7)\\
&= \phi_{11} + 0 + \omega_1 && (8)\\
&= \phi_{11} + \omega_1 && (9)
\end{aligned}$$

$$\begin{aligned}
Var(W_2) &= Var(X_1 + e_2) && (10)\\
&= Cov(X_1 + e_2, X_1 + e_2) && (11)\\
&= Cov(X_1 + e_2, X_1) + Cov(X_1 + e_2, e_2) && (12)\\
&= Cov(X_1 + e_2, X_1) + Cov(X_1, e_2) + Cov(e_2, e_2) && (13)\\
&= Cov(X_1 + e_2, X_1) + 0 + \omega_2 && (14)
\end{aligned}$$

$$= Cov(X_1, X_1) + Cov(e_2, X_1) + \omega_2 \tag{15}$$
$$= \phi_{11} + 0 + \omega_2 \tag{16}$$
$$= \phi_{11} + \omega_2 \tag{17}$$

The variance of the knowledge latent variable is $Var(X_1) = \phi_{11}$. Thus, the reliabilities of these measures are

$$\rho_{W_1} = \frac{\phi_{11}}{\phi_{11} + \omega_1} \tag{18}$$
$$\rho_{W_2} = \frac{\phi_{11}}{\phi_{11} + \omega_2} \tag{19}$$

To obtain their estimates and 95% confidence intervals easily, simply define the parameters within the model and use the function **parameterEstimates()** on the fitted model (see disclosed code). We obtained

- W_1: Estimate of 0.377 and 95% confidence interval of $(0.223, 0.531)$.
- W_2: Estimate of 0.498 and 95% confidence interval of $(0.304, 0.692)$.

(g) We want to test whether the reliabilities of the two knowledge measures are equal. As mentioned, this is equivalent to testing if the measurement error variances are the same,

$$\rho_{W_1} = \rho_{W_2} \tag{20}$$
$$\frac{\phi_{11}}{\phi_{11} + \omega_1} = \frac{\phi_{11}}{\phi_{11} + \omega_2} \tag{21}$$
$$\phi_{11} + \omega_1 = \phi_{11} + \omega_2 \tag{22}$$
$$\omega_1 = \omega_2 \tag{23}$$

We can do this by a W test, where we have 1 row in the **L** matrix with 1 and -1 in the columns corresponding to ω_1 and ω_2 respectively (and 0 in every other column). We would have 1 degree of freedom as we are testing one equality. We have a test statistic of $W = 3.0545096$, $df = 1$ and p-value of 0.0805133. Since p-value is greater than typical significance level of 0.05, there is insufficient evidence to suggest that the measurement error variances for the two knowledge measures are different and consequently insufficient evidence to suggest that the reliabilities of the two knowledge measures are different.

N.B. Note that while we did not do this, an alternate way to carry this out would be to define $\omega_1 - \omega_2$ as a parameter in the model and fit the model. The summary of the model would then effectively display the Z test statistic and p-value for this test.

(h) Taking the assumptions of our model into account, let's also say that $Var(F) = \phi$ (and we are already taking $Var(e_1) = Var(e_2) = \omega$ in this sub-question). Then

$$\begin{aligned}
Cov(D_1, D_2) &= Cov(F + e_1, F + e_2) && (24)\\
&= Cov(F + e_1, F) + Cov(F + e_1, e_2) && (25)\\
&= Cov(F + e_1, F) + Cov(F, e_2) + Cov(e_1, e_2) && (26)\\
&= Cov(F + e_1, F) + 0 + 0 && (27)\\
&= Cov(F, F) + Cov(e_1, F) && (28)\\
&= \phi + 0 && (29)\\
&= \phi && (30)
\end{aligned}$$

And for $i = 1, 2$,

$$\begin{aligned}
Var(D_i) &= Var(F + e_i) && (31)\\
&= Cov(F + e_i, F + e_i) && (32)\\
&= Cov(F + e_i, F) + Cov(F + e_i, e_i) && (33)\\
&= Cov(F + e_i, F) + Cov(F, e_i) + Cov(e_i, e_i) && (34)\\
&= Cov(F + e_i, F) + 0 + \omega && (35)\\
&= Cov(F, F) + Cov(e_i, F) + \omega && (36)\\
&= \phi + 0 + \omega && (37)\\
&= \phi + \omega && (38)
\end{aligned}$$

Now, calculating $Corr(D_1, D_2)$,

$$\begin{aligned}
Corr(D_1, D_2) &= \frac{Cov(D_1, D_2)}{\sqrt{Var(D_1)}\sqrt{Var(D_2)}} && (39)\\
&= \frac{\phi}{\sqrt{\phi + \omega}\sqrt{\phi + \omega}} && (40)
\end{aligned}$$

$$= \frac{\phi}{\phi + \omega} \tag{41}$$

As mentioned, this suggests a sample correlation as an estimate of reliability.

(i) Assuming the measurements of knowledge are equivalent, then based on the findings of the previous sub-question, we can use the sample correlation of the two knowledge measures as an estimate of common reliability. Using the **cor** function to get the sample correlation matrix of all the observable variables as directed, we found this estimate to be 0.44951179 (≈ 0.450). Comparing it to our earlier estimates, we found it to lie somewhere between the earlier estimates (0.377 and 0.498 for W_1 and W_2 respectively).

3

3.1 Problem

Consider the general factor analysis model

$$\mathbf{D}_i = \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i \tag{42}$$

where $\mathbf{\Lambda}$ is a $k \times p$ matrix of factor loadings, the vector of factors $\mathbf{F}_i$ is a $p \times 1$ multivariate normal with expected value zero and covariance matrix $\mathbf{\Phi}$, and $\mathbf{e}_i$ is multivariate normal and independent of $\mathbf{F}_i$, with expected value zero and covariance matrix $\mathbf{\Omega}$. All covariance matrices are positive definite.

(a) How do you know that $\mathbf{D}_i$ is multivariate normal?

(b) Calculate the matrix of covariances between the observable variables $\mathbf{D}_i$ and the underlying factors $\mathbf{F}_i$.

(c) Give the covariance matrix of $\mathbf{D}_i$. Show your work.

(d) Because $\mathbf{\Phi}$ symmetric and positive deifinite, it has a square root matrix that is also symmetric. Using this, show that the parameters of the general factor analysis model are not identifiable.

(e) In an attempt to obtain a model whose parameters can be successfully estimated, let $\mathbf{\Omega}$ be diagonal (errors are uncorrelated) and set $\mathbf{\Phi}$ to the identity matrix (standardized the factors). Show that the parameters of this revised model are still not identifiable. Hint: An orthogonal matrix $\mathbf{R}$ (corresponding to a rotation) is one satisfying $\mathbf{R}\mathbf{R}^T = \mathbf{I}$.

3.2 Solution

(a) $\mathbf{D}_i$ is a linear combination of the multivariate normal vectors $\mathbf{F}_i$ and $\mathbf{e}_i$, making it multivariate normal by the multivariate normal property of linear combination of the multivariate normal vectors being multivariate normal.

(b) We want

$$cov\begin{pmatrix}\mathbf{D}_i\\ \mathbf{F}_i\end{pmatrix} = \left(\begin{array}{c|c} cov(\mathbf{D}_i) & cov(\mathbf{D}_i, \mathbf{F}_i) \\ \hline cov(\mathbf{D}_i, \mathbf{F}_i)^T & cov(\mathbf{F}_i) \end{array}\right) \tag{43}$$

Let's proceed to compute the covariances one by one,

- $cov(\mathbf{D}_i)$:

$$\begin{aligned}
cov(\mathbf{D}_i) &= cov(\mathbf{D}_i, \mathbf{D}_i) && (44)\\
&= cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i) && (45)\\
&= cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \mathbf{\Lambda}\mathbf{F}_i) + cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \mathbf{e}_i) && (46)\\
&= cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \mathbf{\Lambda}\mathbf{F}_i) + cov(\mathbf{\Lambda}\mathbf{F}_i, \mathbf{e}_i) + cov(\mathbf{e}_i, \mathbf{e}_i) && (47)\\
&= cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \mathbf{\Lambda}\mathbf{F}_i) + \mathbf{0} + \mathbf{\Omega} && (48)\\
&= cov(\mathbf{\Lambda}\mathbf{F}_i, \mathbf{\Lambda}\mathbf{F}_i) + cov(\mathbf{e}_i, \mathbf{\Lambda}\mathbf{F}_i) + \mathbf{\Omega} && (49)\\
&= \mathbf{\Lambda} cov(\mathbf{F}_i, \mathbf{F}_i)\mathbf{\Lambda}^T + \mathbf{0} + \mathbf{\Omega} && (50)\\
&= \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Omega} && (51)
\end{aligned}$$

- $cov(\mathbf{D}_i, \mathbf{F}_i)$:

$$\begin{aligned}
cov(\mathbf{D}_i, \mathbf{F}_i) &= cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \mathbf{F}_i) && (52)\\
&= cov(\mathbf{\Lambda}\mathbf{F}_i, \mathbf{F}_i) + cov(\mathbf{e}_i, \mathbf{F}_i) && (53)\\
&= \mathbf{\Lambda} cov(\mathbf{F}_i, \mathbf{F}_i) + \mathbf{0} && (54)\\
&= \mathbf{\Lambda}\mathbf{\Phi} && (55)
\end{aligned}$$

- $cov(\mathbf{F}_i)$: We are already given $cov(\mathbf{F}_i) = \mathbf{\Phi}$.

So we have

$$cov\begin{pmatrix}\mathbf{D}_i \\ \mathbf{F}_i\end{pmatrix} = \left(\begin{array}{c|c} cov(\mathbf{D}_i) & cov(\mathbf{D}_i, \mathbf{F}_i) \\ \hline cov(\mathbf{D}_i, \mathbf{F}_i)^T & cov(\mathbf{F}_i) \end{array}\right) \tag{56}$$

$$= \left(\begin{array}{c|c} \mathbf{\Lambda\Phi\Lambda}^T + \mathbf{\Omega} & \mathbf{\Lambda\Phi} \\ \hline \mathbf{\Phi\Lambda}^T & \mathbf{\Phi} \end{array}\right) \tag{57}$$

If the question only wanted $cov(\mathbf{D}_i, \mathbf{F}_i)$ (this is how I interpreted it when I did this assignment during the course), then we have $cov(\mathbf{D}_i, \mathbf{F}_i) = \mathbf{\Lambda\Phi}$.

(c) We computed $cov(\mathbf{D}_i)$ to be $\mathbf{\Lambda\Phi\Lambda}^T + \mathbf{\Omega}$ in our response to the earlier sub-question (so work done to obtain it is also present there).

(d) We are provided that there is a symmetric square root matrix for the symmetric and positive definite $\mathbf{\Phi}$ (let's call this symmetric square root matrix $\mathbf{\Phi}^{1/2}$). So we have $\mathbf{\Phi} = \mathbf{\Phi}^{1/2}\mathbf{\Phi}^{1/2}$. Now, consider the covariance matrix of the observable data $\mathbf{D}_i$,

$$cov(\mathbf{D}_i) = \mathbf{\Lambda\Phi\Lambda}^T + \mathbf{\Omega} \tag{58}$$

$$= \mathbf{\Lambda\Phi}^{1/2}\mathbf{\Phi}^{1/2}\mathbf{\Lambda}^T + \mathbf{\Omega} \tag{59}$$

$$= (\mathbf{\Lambda\Phi}^{1/2})(\mathbf{\Lambda\Phi}^{1/2})^T + \mathbf{\Omega} \tag{60}$$

$$= (\mathbf{\Lambda\Phi}^{1/2})\mathbf{I}(\mathbf{\Lambda\Phi}^{1/2})^T + \mathbf{\Omega} \tag{61}$$

$$= \mathbf{\Lambda}_2\mathbf{I}\mathbf{\Lambda}_2^T + \mathbf{\Omega} \tag{62}$$

With this we note that both sets of parameters $(\mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Omega})$ and $(\mathbf{\Lambda}_2, \mathbf{I}, \mathbf{\Omega})$ (here $\mathbf{\Lambda}_2 = \mathbf{\Lambda\Phi}^{1/2}$) produce the same observable data covariance matrix (and we would not be able to determine which set of parameters generates the observable data covariance matrix from the data). As a result the parameters of the general factor analysis model are not identifiable.

(e) Provided that we let $\mathbf{\Omega}$ be diagonal (errors are uncorrelated) and set $\mathbf{\Phi}$ to the identity matrix (standardized the factors), then the observable data covariance matrix is $\mathbf{\Lambda I\Lambda}^T + \mathbf{\Omega}$. Making use of the hint, take an orthogonal matrix $\mathbf{R}$ (i.e. this is a matrix that satisfies $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ as per hint). Consider then,

$$
\begin{aligned}
cov(\mathbf{D}_i) &= \mathbf{\Lambda}\mathbf{I}\mathbf{\Lambda}^T + \mathbf{\Omega} && (63)\\
&= \mathbf{\Lambda}\mathbf{R}\mathbf{R}^T\mathbf{\Lambda}^T + \mathbf{\Omega} && (64)\\
&= (\mathbf{\Lambda}\mathbf{R})(\mathbf{\Lambda}\mathbf{R})^T + \mathbf{\Omega} && (65)\\
&= \mathbf{\Lambda}_3\mathbf{\Lambda}_3^T + \mathbf{\Omega} && (66)\\
& && (67)
\end{aligned}
$$

Again we note that both sets of parameters $(\mathbf{\Lambda}, \mathbf{\Omega})$ and $(\mathbf{\Lambda}_3, \mathbf{\Omega})$ (here $\mathbf{\Lambda}_3 = \mathbf{\Lambda}\mathbf{R}$) produce the same observable data covariance matrix (and we would not be able to determine which set of parameters generates the observable data covariance matrix from the data). As a result the parameters of this revised model are still not identifiable.

4

4.1 Problem

In this factor analysis model, the observed variables are *not* standardized, and the factor loading for D_1 is set equal to one. Let

$$
\begin{aligned}
D_1 &= F + e_1 && (68)\\
D_2 &= \lambda_2 F + e_2 && (69)\\
D_3 &= \lambda_3 F + e_3, && (70)
\end{aligned}
$$

where $F \sim N(0, \phi)$, e_1, e_2 and e_3 are normal and independent of F and each other with expected value zero, $Var(e_1) = \omega_1$, $Var(e_2) = \omega_2$, $Var(e_3) = \omega_3$, and λ_2 and λ_3 are nonzero constants.

(a) Calculate the variance-covariance matrix of observed variables.

(b) Are the model parameters identifiable? Answer Yes or No and prove your answer.

4.2 Solution

(a) Let

$$cov\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \hline & \sigma_{22} & \sigma_{23} \\ \hline & & \sigma_{33} \end{array}\right) = \mathbf{\Sigma}. \tag{71}$$

Let's proceed to compute the variances and covariances one by one

- σ_{11}:

$$\begin{aligned} \sigma_{11} &= Cov(D_1, D_1) & (72)\\ &= Cov(F + e_1, F + e_1) & (73)\\ &= Cov(F + e_1, F) + Cov(F + e_1, e_1) & (74)\\ &= Cov(F + e_1, F) + Cov(F, e_1) + Cov(e_1, e_1) & (75)\\ &= Cov(F + e_1, F) + 0 + \omega_1 & (76)\\ &= Cov(F, F) + Cov(e_1, F) + \omega_1 & (77)\\ &= \phi + 0 + \omega_1 & (78)\\ &= \phi + \omega_1 & (79) \end{aligned}$$

- σ_{1j} for $j = 2, 3$:

$$\begin{aligned} \sigma_{1j} &= Cov(D_1, D_j) & (80)\\ &= Cov(F + e_1, \lambda_j F + e_j) & (81)\\ &= Cov(F + e_1, \lambda_j F) + Cov(F + e_1, e_j) & (82)\\ &= Cov(F + e_1, \lambda_j F) + 0 & (83)\\ &= Cov(F, \lambda_j F) + Cov(e_1, \lambda_j F) & (84)\\ &= \lambda_j Cov(F, F) + 0 & (85)\\ &= \lambda_j \phi & (86) \end{aligned}$$

- σ_{jj} for $j = 2, 3$:

$$\begin{aligned} \sigma_{jj} &= Cov(D_j, D_j) & (87)\\ &= Cov(\lambda_j F + e_j, \lambda_j F + e_j) & (88)\\ &= Cov(\lambda_j F + e_j, \lambda_j F) + Cov(\lambda_j F + e_j, e_j) & (89) \end{aligned}$$

$$
\begin{aligned}
&= Cov(\lambda_j F + e_j, \lambda_j F) + Cov(\lambda_j F, e_j) + Cov(e_j, e_j) && (90)\\
&= Cov(\lambda_j F + e_j, \lambda_j F) + 0 + \omega_j && (91)\\
&= Cov(\lambda_j F, \lambda_j F) + Cov(e_j, \lambda_j F) + \omega_j && (92)\\
&= \lambda_j^2 Cov(F, F) + 0 + \omega_j && (93)\\
&= \lambda_j^2 \phi + \omega_j && (94)
\end{aligned}
$$

- σ_{23}:

$$
\begin{aligned}
\sigma_{23} &= Cov(D_2, D_3) && (95)\\
&= Cov(\lambda_2 F + e_2, \lambda_3 F + e_3) && (96)\\
&= Cov(\lambda_2 F + e_2, \lambda_3 F) + Cov(\lambda_2 F + e_2, e_3) && (97)\\
&= Cov(\lambda_2 F + e_2, \lambda_3 F) + 0 && (98)\\
&= Cov(\lambda_2 F, \lambda_3 F) + Cov(e_2, \lambda_3 F) && (99)\\
&= \lambda_2 \lambda_3 Cov(F, F) + 0 && (100)\\
&= \lambda_2 \lambda_3 \phi && (101)
\end{aligned}
$$

So we have

$$
\begin{aligned}
cov \begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} &= \left(\begin{array}{c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \hline & \sigma_{22} & \sigma_{23} \\ \hline & & \sigma_{33} \end{array} \right) && (102)\\
&= \left(\begin{array}{c|c|c} \phi + \omega_1 & \lambda_2 \phi & \lambda_3 \phi \\ \hline & \lambda_2^2 \phi + \omega_2 & \lambda_2 \lambda_3 \phi \\ \hline & & \lambda_3^2 \phi + \omega_3 \end{array} \right) && (103)
\end{aligned}
$$

(b) Let's attempt to solve for the parameters

- λ_2:

$$
\begin{aligned}
\frac{\lambda_2 \lambda_3 \phi}{\lambda_3 \phi} &= \frac{\sigma_{23}}{\sigma_{13}} && (104)\\
\lambda_2 &= \frac{\sigma_{23}}{\sigma_{13}} && (105)
\end{aligned}
$$

- λ_3:

$$
\begin{aligned}
\frac{\lambda_2 \lambda_3 \phi}{\lambda_2 \phi} &= \frac{\sigma_{23}}{\sigma_{12}} && (106)\\
\lambda_3 &= \frac{\sigma_{23}}{\sigma_{12}} && (107)
\end{aligned}
$$

- ϕ:

$$\frac{\lambda_2\phi\lambda_3\phi}{\lambda_2\lambda_3\phi} = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \tag{108}$$

$$\frac{\lambda_2\lambda_3\phi^2}{\lambda_2\lambda_3\phi} = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \tag{109}$$

$$\phi = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \tag{110}$$

- ω_1:

$$\phi + \omega_1 = \sigma_{11} \tag{111}$$

$$\omega_1 = \sigma_{11} - \phi \tag{112}$$

- ω_2:

$$\lambda_2^2\phi + \omega_2 = \sigma_{22} \tag{113}$$

$$\omega_2 = \sigma_{22} - \lambda_2^2\phi \tag{114}$$

- ω_3:

$$\lambda_3^2\phi + \omega_3 = \sigma_{33} \tag{115}$$

$$\omega_3 = \sigma_{33} - \lambda_3^2\phi \tag{116}$$

Assuming the variance parameters cannot be 0, we note that the above solutions apply if the factor loadings are non-zero. Thus provided the factor loadings are non-zero, we can solve for all the parameters from the covariance structure equations, consequently making all model parameters identifiable.

We can identify if either of the factor loadings is zero. This is so since we can test for $\lambda_j = 0$ by testing for $\sigma_{1j} = 0$ for $j = 2, 3$ (note that the variance parameters cannot be 0).

If both factor loadings are 0, then we can identify ω_2 and ω_3 ($\omega_2 = \sigma_{22}$ and $\omega_3 = \sigma_{33}$ in this case), but we can't identify ϕ and ω_1 (since we are left with 1 covariance structure equation for these two parameters).

If only one of the factor loadings is 0 (for example, let's take $\lambda_2 = 0$), then the factor loading that is 0 can be identified as discussed earlier and its corresponding measurement error variance can be identified (so in our example we would have $\omega_2 = \sigma_{22}$). However, we would then be left with 4 parameters and 3 remaining covariance structure equations, so the remaining 4 parameters can't be identified.

After going through all these, I just noticed that we are assuming the factor loadings are non-zero constants. So, if we are assuming this and the factor loading value of zero for either of the factor loadings isn't in the parameter space, then the non-zero case of the factor loadings in our investigation applies. In that case, **Yes**, the model parameters are identifiable.

5

5.1 Problem

We now extend the preceding model by adding another factor. Let

$$\begin{aligned} D_1 &= F_1 + e_1 && (117)\\ D_2 &= \lambda_2 F_1 + e_2 && (118)\\ D_3 &= \lambda_3 F_1 + e_3 && (119)\\ D_4 &= F_2 + e_4 && (120)\\ D_5 &= \lambda_5 F_2 + e_5 && (121)\\ D_6 &= \lambda_6 F_2 + e_6, && (122) \end{aligned}$$

where all expected values are zero, $Var(e_i) = \omega_i$ for $i = 1, ..., 6$,

$$cov \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}, \tag{123}$$

and λ_2, λ_3, λ_5 and λ_6 are nonzero constants.

(a) Give the covariance matrix of the observable variables. Show the necessary work. A lot of the work has already been done in Question 4.

(b) Are the model parameters identifiable? Answer Yes or No and prove your answer.

5.2 Solution

(a) Let

$$
cov\begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \\ D_6 \end{pmatrix} = \left(\begin{array}{c|c|c|c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \hline & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \hline & & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \hline & & & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \hline & & & & \sigma_{55} & \sigma_{56} \\ \hline & & & & & \sigma_{66} \end{array}\right) = \mathbf{\Sigma}. \tag{124}
$$

Let's proceed to compute the variances and covariances one by one

- We note that for the following component of the observable data covariance matrix

$$
cov\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \hline & \sigma_{22} & \sigma_{23} \\ \hline & & \sigma_{33} \end{array}\right). \tag{125}
$$

  We can simply use the work we have done in the previous question as hinted by this sub-question. Specifically assuming that the errors are independent of the factors and each other (since we are being asked to borrow from our work in the previous question, I am assuming the same assumptions hold), this component of the observable data covariance matrix corresponds to exactly the same model (same equations and assumptions) we investigated in the previous question. So we can simply use our computations from the previous question,

$$
cov\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \hline & \sigma_{22} & \sigma_{23} \\ \hline & & \sigma_{33} \end{array}\right) \tag{126}
$$

$$
= \left(\begin{array}{c|c|c} \phi_{11} + \omega_1 & \lambda_2\phi_{11} & \lambda_3\phi_{11} \\ \hline & \lambda_2^2\phi_{11} + \omega_2 & \lambda_2\lambda_3\phi_{11} \\ \hline & & \lambda_3^2\phi_{11} + \omega_3 \end{array}\right) \tag{127}
$$

  N.B. Since the factor variance is now $Var(F_1) = \phi_{11}$, we had to make the substitution $\phi = \phi_{11}$ to our solution from the previous question. The other parameters seems to have the same labels so they remain unchanged.

- We also note that the previous point also applies to the following component

$$cov\begin{pmatrix} D_4 \\ D_5 \\ D_6 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \hline & \sigma_{55} & \sigma_{56} \\ \hline & & \sigma_{66} \end{array}\right). \tag{128}$$

So we use the same solution, making sure to do the necessary substitutions (i.e. $\lambda_2 = \lambda_5$, $\lambda_3 = \lambda_6$, $\phi = \phi_{22}$, $\omega_1 = \omega_4$, $\omega_2 = \omega_5$, $\omega_3 = \omega_6$),

$$cov\begin{pmatrix} D_4 \\ D_5 \\ D_6 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \hline & \sigma_{55} & \sigma_{56} \\ \hline & & \sigma_{66} \end{array}\right) \tag{129}$$

$$= \left(\begin{array}{c|c|c} \phi_{22} + \omega_4 & \lambda_5\phi_{22} & \lambda_6\phi_{22} \\ \hline & \lambda_5^2\phi_{22} + \omega_5 & \lambda_5\lambda_6\phi_{22} \\ \hline & & \lambda_6^2\phi_{22} + \omega_6 \end{array}\right) \tag{130}$$

- All that's left is to do the computations for the remaining component

$$cov\left(\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix}, \begin{pmatrix} D_4 \\ D_5 \\ D_6 \end{pmatrix}\right) = \left(\begin{array}{c|c|c} \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \hline \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \hline \sigma_{34} & \sigma_{35} & \sigma_{36} \end{array}\right). \tag{131}$$

Let's proceed to compute these remaining covariances one by one

  - σ_{14}:

$$\sigma_{14} = Cov(D_1, D_4) \tag{132}$$
$$= Cov(F_1 + e_1, F_2 + e_4) \tag{133}$$
$$= Cov(F_1 + e_1, F_2) + Cov(F_1 + e_1, e_4) \tag{134}$$
$$= Cov(F_1 + e_1, F_2) + 0 \tag{135}$$
$$= Cov(F_1, F_2) + Cov(e_1, F_2) \tag{136}$$
$$= \phi_{12} + 0 \tag{137}$$
$$= \phi_{12} \tag{138}$$

- σ_{1j} for $j = 5, 6$:

$$\sigma_{1j} = Cov(D_1, D_j) \tag{139}$$
$$= Cov(F_1 + e_1, \lambda_j F_2 + e_j) \tag{140}$$
$$= Cov(F_1 + e_1, \lambda_j F_2) + Cov(F_1 + e_1, e_j) \tag{141}$$
$$= Cov(F_1 + e_1, \lambda_j F_2) + Cov(F_1 + e_1, e_j) \tag{142}$$
$$= Cov(F_1 + e_1, \lambda_j F_2) + 0 \tag{143}$$
$$= Cov(F_1, \lambda_j F_2) + Cov(e_1, \lambda_j F_2) \tag{144}$$
$$= \lambda_j Cov(F_1, F_2) + 0 \tag{145}$$
$$= \lambda_j \phi_{12} \tag{146}$$

- σ_{i4} for $i = 2, 3$:

$$\sigma_{i4} = Cov(D_i, D_4) \tag{147}$$
$$= Cov(\lambda_i F_1 + e_i, F_2 + e_4) \tag{148}$$
$$= Cov(\lambda_i F_1 + e_i, F_2) + Cov(\lambda_i F_1 + e_i, e_4) \tag{149}$$
$$= Cov(\lambda_i F_1 + e_i, F_2) + 0 \tag{150}$$
$$= Cov(\lambda_i F_1, F_2) + Cov(e_i, F_2) \tag{151}$$
$$= \lambda_i Cov(F_1, F_2) + 0 \tag{152}$$
$$= \lambda_i \phi_{12} \tag{153}$$

- σ_{ij} for $i = 2, 3$ and $j = 5, 6$:

$$\sigma_{ij} = Cov(D_i, D_j) \tag{154}$$
$$= Cov(\lambda_i F_1 + e_i, \lambda_j F_2 + e_j) \tag{155}$$
$$= Cov(\lambda_i F_1 + e_i, \lambda_j F_2) + Cov(\lambda_i F_1 + e_i, e_j) \tag{156}$$
$$= Cov(\lambda_i F_1 + e_i, \lambda_j F_2) + 0 \tag{157}$$
$$= Cov(\lambda_i F_1, \lambda_j F_2) + Cov(e_i, \lambda_j F_2) \tag{158}$$
$$= \lambda_i \lambda_j Cov(F_1, F_2) + 0 \tag{159}$$
$$= \lambda_i \lambda_j \phi_{12} \tag{160}$$

So we have

$$cov \begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \\ D_5 \\ D_6 \end{pmatrix} = \left(\begin{array}{c|c|c|c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \hline & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \hline & & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \hline & & & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \hline & & & & \sigma_{55} & \sigma_{56} \\ \hline & & & & & \sigma_{66} \end{array}\right) \tag{161}$$

$$= \left(\begin{array}{c|c|c|c|c|c} \phi_{11}+\omega_1 & \lambda_2\phi_{11} & \lambda_3\phi_{11} & \phi_{12} & \lambda_5\phi_{12} & \lambda_6\phi_{12} \\ \hline & \lambda_2^2\phi_{11}+\omega_2 & \lambda_2\lambda_3\phi_{11} & \lambda_2\phi_{12} & \lambda_2\lambda_5\phi_{12} & \lambda_2\lambda_6\phi_{12} \\ \hline & & \lambda_3^2\phi_{11}+\omega_3 & \lambda_3\phi_{12} & \lambda_3\lambda_5\phi_{12} & \lambda_3\lambda_6\phi_{12} \\ \hline & & & \phi_{22}+\omega_4 & \lambda_5\phi_{22} & \lambda_6\phi_{22} \\ \hline & & & & \lambda_5^2\phi_{22}+\omega_5 & \lambda_5\lambda_6\phi_{22} \\ \hline & & & & & \lambda_6^2\phi_{22}+\omega_6 \end{array}\right) \tag{162}$$

(b) Let's try to solve for the parameters to see if they're identifiable.

- As with the previous sub-question, we note that for the following component

$$cov \begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \hline & \sigma_{22} & \sigma_{23} \\ \hline & & \sigma_{33} \end{array}\right) \tag{163}$$

$$= \left(\begin{array}{c|c|c} \phi_{11}+\omega_1 & \lambda_2\phi_{11} & \lambda_3\phi_{11} \\ \hline & \lambda_2^2\phi_{11}+\omega_2 & \lambda_2\lambda_3\phi_{11} \\ \hline & & \lambda_3^2\phi_{11}+\omega_3 \end{array}\right), \tag{164}$$

  we already have solved for the parameters within this component in the previous question,

  - $\lambda_2 = \frac{\sigma_{23}}{\sigma_{13}}$
  - $\lambda_3 = \frac{\sigma_{23}}{\sigma_{12}}$
  - $\phi_{11} = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}$
  - $\omega_1 = \sigma_{11} - \phi_{11}$
  - $\omega_2 = \sigma_{22} - \lambda_2^2\phi_{11}$
  - $\omega_3 = \sigma_{33} - \lambda_3^2\phi_{11}$

N.B. Since the factor variance is now $Var(F_1) = \phi_{11}$, we had to make the substitution $\phi = \phi_{11}$ to our solution from the previous question. The other parameters seems to have the same labels so they remain unchanged.

- We also note that the previous point also applies to the following component

$$cov \begin{pmatrix} D_4 \\ D_5 \\ D_6 \end{pmatrix} = \left(\begin{array}{c|c|c} \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \hline & \sigma_{55} & \sigma_{56} \\ \hline & & \sigma_{66} \end{array}\right) \tag{165}$$

$$= \left(\begin{array}{c|c|c} \phi_{22} + \omega_4 & \lambda_5 \phi_{22} & \lambda_6 \phi_{22} \\ \hline & \lambda_5^2 \phi_{22} + \omega_5 & \lambda_5 \lambda_6 \phi_{22} \\ \hline & & \lambda_6^2 \phi_{22} + \omega_6 \end{array}\right), \tag{166}$$

that is with the necessary substitutions. Specifically, $\lambda_2 = \lambda_5$, $\lambda_3 = \lambda_6$, $\phi = \phi_{22}$, $\omega_1 = \omega_4$, $\omega_2 = \omega_5$, $\omega_3 = \omega_6$, $\sigma_{11} = \sigma_{44}$, $\sigma_{12} = \sigma_{45}$, $\sigma_{13} = \sigma_{46}$, $\sigma_{22} = \sigma_{55}$, $\sigma_{23} = \sigma_{56}$, $\sigma_{33} = \sigma_{66}$. So we have,

  - $\lambda_5 = \frac{\sigma_{56}}{\sigma_{46}}$
  - $\lambda_6 = \frac{\sigma_{56}}{\sigma_{45}}$
  - $\phi_{22} = \frac{\sigma_{45}\sigma_{46}}{\sigma_{56}}$
  - $\omega_4 = \sigma_{44} - \phi_{22}$
  - $\omega_5 = \sigma_{55} - \lambda_5^2 \phi_{22}$
  - $\omega_6 = \sigma_{66} - \lambda_6^2 \phi_{22}$

- All that remains is ϕ_{12}, for which we see right away $\phi_{12} = \sigma_{14}$.

Thus, we can solve for all the parameters from the covariance structure equations. Consequently, **Yes**, all model parameters are identifiable.

6

6.1 Problem

Let's add a third factor to the model of Question 5. That is we add

$$D_7 = F_3 + e_7 \tag{167}$$

$$D_8 = \lambda_8 F_3 + e_8 \tag{168}$$
$$D_9 = \lambda_9 F_3 + e_9 \tag{169}$$

and

$$cov\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix}, \tag{170}$$

with $\lambda_8 \neq 0$, $\lambda_9 \neq 0$ and so on. Are the model parameters identifiable? You don't have to do any calculations if you see the pattern.

Bring a printout with your R input and output to the quiz. Please remember that while the questions may appear in comment statements, answers and interpretation may not, except for numerical answers generated by R.

6.2 Solution

As mentioned by the question, we observe an emerging pattern from our responses to **4(b)** and **5(b)**. Specifically

- We can solve for λ_2, λ_3, ϕ_{11}, ω_1, ω_2 and ω_3 from the component $cov\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix}$ of the observable data covariance matrix (the situation is exactly the same (with some label substitutions) as that in Question 4, where we show the work on how to solve for these parameters).

- We can solve for λ_5, λ_6, ϕ_{22}, ω_4, ω_5 and ω_6 from the component $cov\begin{pmatrix} D_4 \\ D_5 \\ D_6 \end{pmatrix}$ of the observable data covariance matrix (the situation is exactly the same (with some label substitutions) as that in Question 4, where we show the work on how to solve for these parameters).

- We can solve for λ_8, λ_9, ϕ_{33}, ω_7, ω_8 and ω_9 from the component $cov\begin{pmatrix} D_7 \\ D_8 \\ D_9 \end{pmatrix}$ of the observable data covariance matrix (the situation is exactly the same (with some label substitutions) as that in Question 4, where we show the work on how to solve for these parameters).

- From the previous question we have $\phi_{12} = \sigma_{14}$ (we have done the work to obtain this there). This applies in this question as well, since the situation is the same.

- We can extrapolate from how we obtained the result in the previous point that

$$\begin{aligned}
\sigma_{17} &= cov(D_1, D_7) && (171)\\
&= cov(F_1 + e_1, F_3 + e_7) && (172)\\
&= cov(F_1 + e_1, F_3) + cov(F_1 + e_1, e_7) && (173)\\
&= cov(F_1 + e_1, F_3) + 0 && (174)\\
&= cov(F_1, F_3) + cov(e_1, F_3) && (175)\\
&= \phi_{13} + 0 && (176)\\
&= \phi_{13} && (177)
\end{aligned}$$

  That is we have $\phi_{13} = \sigma_{17}$.

- Likewise, similar to the previous point,

$$\begin{aligned}
\sigma_{47} &= cov(D_4, D_7) && (178)\\
&= cov(F_2 + e_4, F_3 + e_7) && (179)\\
&= cov(F_2 + e_4, F_3) + cov(F_2 + e_4, e_7) && (180)\\
&= cov(F_2 + e_4, F_3) + 0 && (181)\\
&= cov(F_2, F_3) + cov(e_4, F_3) && (182)\\
&= \phi_{23} + 0 && (183)\\
&= \phi_{23} && (184)
\end{aligned}$$

  That is we have $\phi_{23} = \sigma_{47}$.

Thus, we can solve for all the parameters from the covariance structure equations. Consequently, **Yes**, all model parameters are identifiable.