NAME (PRINT): ____

Last/Surname First /Given Name

STUDENT #: ______________________ **SIGNATURE:** ________________

UNIVERSITY OF TORONTO MISSISSAUGA
APRIL 2011 FINAL EXAMINATION
STA431H5S
Structural Equation Models
Jerry Brunner
Duration - 3 hours
Aids: Statistical Calculators; Reference Sheet (Supplied)

The University of Toronto Mississauga and you, as a student, share a commitment to academic integrity. You are reminded that you may be charged with an academic offence for possessing any unauthorized aids during the writing of an exam, including but not limited to any electronic devices with storage, such as cell phones, pagers, personal digital assistants (PDAs), iPods, and MP3 players. Unauthorized calculators and notes are also not permitted. Do not have any of these items in your possession in the area of your desk. Please turn the electronics off and put all unauthorized aids with your belongings at the front of the room before the examination begins. If any of these items are kept with you during the writing of your exam, you may be charged with an academic offence. A typical penalty may cause you to fail the course.

Please note, you ***CANNOT*** *petition to* ***re-write*** *an examination once the exam has begun.*

Qn. #	Value	Score
1	10	
2	27	
3	23	
4	12	
5	10	
6	18	
Total = 100 Points		

10 points

1. Let F be a latent variable. Consider two equivalent measurements with uncorrelated measurement error:

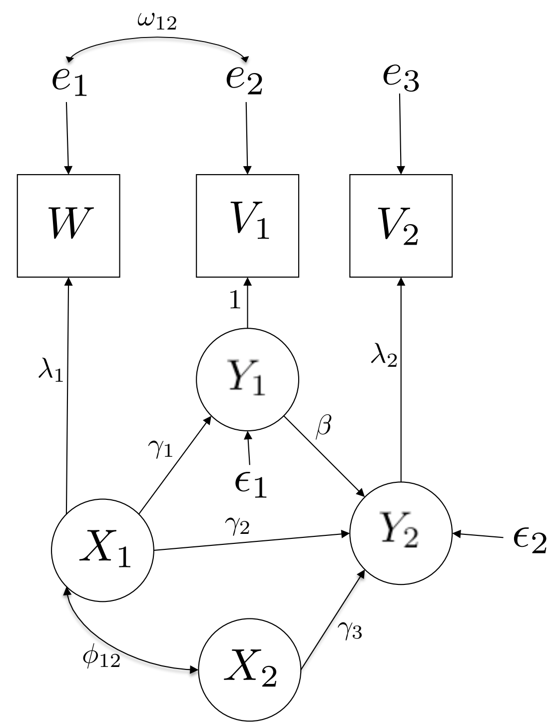
$$
\begin{aligned}
D_1 &= \nu + F + e_1 \\
D_2 &= \nu + F + e_2,
\end{aligned}
$$

where $E(F) = \mu$, $Var(F) = \phi$, $E(e_1) = E(e_2) = 0$, $Var(e_1) = Var(e_2) = \omega$, and F, e_1 and e_2 are all independent. The observable variables in this model are D_1 and D_2. The *reliability* of an observable measurement is defined as the squared correlation between the true (latent) measurement and the observable measurement. Show that the reliability of D_1 equals the correlation between D_1 and D_2. Hint: both quantities equal a familiar formula.

continued on page 3

27 points

2. Consider the following path diagram.



(a) Write down the model equations in scalar form.

continued on page 4

(b) Refer to the general two-stage structural equation model on the formula sheet. Write down the following, using symbols from the path diagram. This part is intended to help you use matrices with the correct dimensions.

i. $\mathbf{X}$

ii. $\mathbf{Y}$

iii. $\mathbf{F}$

iv. $\mathbf{D}$

(c) Give the matrix $\boldsymbol{\beta}$, using symbols from the path diagram. *Make sure it has the correct dimensions.*

continued on page 5

(d) Give the matrix $\mathbf{\Gamma}$, using symbols from the path diagram. Make sure it has the correct dimensions.

(e) Give the matrix $\mathbf{\Lambda}$, using symbols from the path diagram. Make sure it has the correct dimensions.

(f) Give the matrix $\mathbf{\Phi}_{11}$, using symbols from the path diagram. Make sure it has the correct dimensions. Not all symbols in $\mathbf{\Phi}_{11}$ appear on the path diagram.

(g) Give the matrix $\mathbf{\Psi}$. Make sure it has the correct dimensions. The symbols in $\mathbf{\Psi}$ do not appear on the path diagram. Use standard notation.

(h) Give the matrix $\mathbf{\Omega}$, using symbols from the path diagram. Make sure it has the correct dimensions. Not all symbols in $\mathbf{\Omega}$ appear on the path diagram.

continued on page 6

23 points

3. For the measurement part of the general two-stage structural equation model (see formula sheet), suppose $\mathbf{\Phi}$ is the identity and $\mathbf{\Omega}$ is diagonal. In addition, suppose that the observable variables have been standardized, and we write $\mathbf{Z}$ in place of $\mathbf{D}$. This yields the classical exploratory factor analysis model. In your calculations, assume $\mathbf{F}$ is $p \times 1$ and $\mathbf{Z}$ is $k \times 1$.

   (a) Show that factor loadings are actually correlations by calculating $Corr(Z_i, F_j)$. This is a *scalar* (not matrix) calculation. Show your work. *Circle your final answer.*

   (b) Show that $\omega_{i,i} = Var(e_i)$ is a function of the $\lambda_{i,j}$ quantities. This is a *scalar* (not matrix) calculation. Show your work. *Circle your final answer.*

   (c) What are the *unknown* parameters in this classical factor analysis model? The answer is a list of one or more matrices.

continued on page 7

(d) Prove that the parameters of the the classical exploratory factor analysis model (the model you are working with in this question) are not identifiable. This is the case even if there are enough observable variables so that the counting rule is satisfied, so don't try to use the counting rule. Show your work.

continued on page 8

12 points

4. let

$$\begin{aligned} D_1 &= \lambda_1 F + e_1 \\ D_2 &= \lambda_2 F + e_2 \\ D_3 &= \lambda_3 F + e_3, \end{aligned}$$

where all expected values are zero, $V(e_i) = \omega_i$ for $i = 1, 2, 3$, $V(F) = \phi$, the error terms are independent of one another and of the factor, $\lambda_1 > 0$, $\lambda_2 \neq 0$ and $\lambda_3 \neq 0$. The parameters of the model are not identifiable (you don't have to show this), but you can work with it anyway; *do not re-parameterize*!

(a) Give the covariance matrix $\boldsymbol{\Sigma}$. This is a scalar calculation.

(b) Show that the ratio of factor loadings λ_1/λ_2 is identifiable.

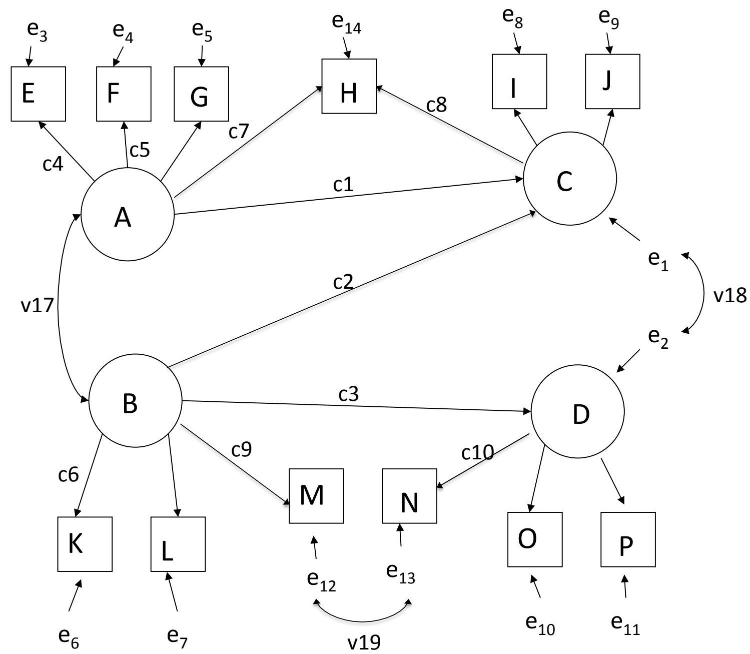
continued on page 9

(c) Show that the parameter ω_1 is identifiable. You have more room on this page than you need.

continued on page 10

10 points

5. The parameters of the model represented by the following path diagram are identifiable. Briefly explain why. Make specific reference to the identifiability rules on the reference sheet.



continued on page 11

18 points

6. The *Poverty Data* contain information from a sample of 97 countries. The variables include Live birth rate per 1,000 of population, Death rate per 1,000 of population, Infant deaths per 1,000 of population under 1 year old, Life expectancy at birth for males, Life expectancy at birth for females, and Gross National Product per capita in U.S. dollars. Please answer these questions based on the SAS program and list file starting on Page 12.

   Note: The printout shows two runs of `proc calis`. Each one fits a different model.

   (a) What is the parameter $\boldsymbol{\theta}$ for the factor analysis model? That's the first run of `proc calis`. *Give your answer in the form of a list of names from the SAS job.*

   (b) Does the model fit the data adequately? Answer Yes or No, and back up your answer with two numbers from the printout: The value of a test statistic, and a p-value.

   (c) What is the maximum likelihood estimate of the correlation between factors? The answer is a single number from the printout.

   (d) Clearly, the factor loadings are not identifiable for points in the parameter space where $\phi_{12} = 0$ In your view, is this likely to be a problem in the present case? Answer "Yes, a problem," or "No problem" and briefly explain. Does the printout support your position?

   (e) Now consider the second run of `proc calis` – the latent regression model. Does this model fit the data adequately? Answer Yes or No. Is the fit better than the factor analysis model, worse than the factor analysis model, or is it the same? Write one of these alternatives and *circle it.*

   (f) According to the regression model, wealth causes health if there is any connection between the two. How do you know *from the printout* that more money translates into better health (and not worse health)? Cite a particular number from the printout.

That's the end of the questions. The rest of this exam script is computer printout.

Total Marks = 100 points

continued on page 12

```
/********************* finalpoverty2011B.sas ************************/
options linesize=79 noovp formdlim='-';
title 'UN Poverty Data: STA431 S 2011 Final Exam';

data misery;
     infile 'poverty.data';
     input birthrate deathrate infmort lifexM lifexF gnp group country $;
     gnp1000 = gnp/1000; /* In thousands of dollars */
     lifex = (lifexM+lifexF)/2;

proc calis cov;
     title2 'Factor analysis model standardizing the factors';
     var lifex infmort gnp1000 birthrate;
     lineqs
          lifex     = lambda1 Fhealth + e1,
          infmort   = lambda2 Fhealth + e2,
          gnp1000   = lambda3 Fwealth + e3,
          birthrate = lambda4 Fwealth + e4;
     std
          Fhealth = 1, Fwealth=1,
          e1-e4 = 4 * omega: ;
     cov  Fhealth Fwealth = phi12;
     bounds 0.0 <  omega1-omega4;

proc calis cov;
     title2 'Latent regression model setting loadings to one';
     var lifex infmort gnp1000 birthrate;
     lineqs
          Fhealth   = gamma Fwealth + epsilon,
          lifex     =         Fhealth + e1,
          infmort   = lambda2 Fhealth + e2,
          gnp1000   =         Fwealth + e3,
          birthrate = lambda4 Fwealth + e4;
     std
          Fwealth=phi, epsilon = psi,
          e1-e4 = 4 * omega: ;
     bounds 0.0 < phi psi omega1-omega4;
```

continued on page 13

Parts of finalpoverty2011B.lst

```
-------------------------------------------------------------------------------
                   UN Poverty Data: STA431 S 2011 Final Exam                  1
                Factor analysis model standardizing the factors

                          The 4 Endogenous Variables

    Manifest        lifex      infmort    gnp1000    birthrate
    Latent


                          The 6 Exogenous Variables

    Manifest
    Latent          Fhealth    Fwealth
    Error           e1         e2         e3         e4

-------------------------------------------------------------------------------
                   UN Poverty Data: STA431 S 2011 Final Exam                  5
                Factor analysis model standardizing the factors

                                                                       Actual
                                                   Max Abs               Over
        Rest    Func      Act    Objective   Obj Fun Gradient             Pred
 Iter   arts   Calls      Con     Function    Change  Element  Lambda   Change

    1      0       2        0      0.01215   0.00548   0.0285       0    1.099
    2      0       3        0      0.01207  0.000077 0.000084       0    1.024
    3      0       4        0      0.01207  1.067E-6 0.000012       0    1.024
    4      0       5        0      0.01207    1.6E-8 1.267E-6       0    1.023

                             Optimization Results

Iterations                         4  Function Calls                   6
Jacobian Calls                     5  Active Constraints               0
Objective Function      0.0120694315  Max Abs Gradient Element  1.2666839E-6
Lambda                             0  Actual Over Pred Change   1.0233020931
Radius                  0.0010158686

ABSGCONV convergence criterion satisfied.
```

continued on page 14

```
-------------------------------------------------------------------------------
                 UN Poverty Data: STA431 S 2011 Final Exam                   6
              Factor analysis model standardizing the factors

                            The CALIS Procedure
         Covariance Structure Analysis: Maximum Likelihood Estimation

         Fit Function                                               0.0121
         Goodness of Fit Index (GFI)                                0.9940
         GFI Adjusted for Degrees of Freedom (AGFI)                 0.9404
         Root Mean Square Residual (RMR)                            2.5734
         Standardized Root Mean Square Residual (SRMR)              0.0067
         Parsimonious GFI (Mulaik, 1989)                            0.1657
         Chi-Square                                                 1.0862
         Chi-Square DF                                                   1
         Pr > Chi-Square                                            0.2973
         Independence Model Chi-Square                              399.83
         Independence Model Chi-Square DF                                6

-------------------------------------------------------------------------------
                 UN Poverty Data: STA431 S 2011 Final Exam                   7
              Factor analysis model standardizing the factors

                            The CALIS Procedure
         Covariance Structure Analysis: Maximum Likelihood Estimation

                   Manifest Variable Equations with Estimates


              lifex     =  10.3124*Fhealth   +  1.0000 e1
              Std Err       0.7841 lambda1
              t Value      13.1522
              infmort   = -44.2690*Fhealth   +  1.0000 e2
              Std Err       3.6165 lambda2
              t Value     -12.2408
              gnp1000   =   5.4842*Fwealth   +  1.0000 e3
              Std Err       0.7664 lambda3
              t Value       7.1560
              birthrate = -12.7179*Fwealth   +  1.0000 e4
              Std Err       1.1530 lambda4
              t Value     -11.0302


                        Variances of Exogenous Variables

                                                Standard
          Variable  Parameter      Estimate        Error    t Value

          Fhealth                   1.00000
          Fwealth                   1.00000
          e1        omega1          1.44765      1.84498       0.78
          e2        omega2        184.16184     43.51831       4.23
          e3        omega3         35.43109      5.57604       6.35
          e4        omega4         25.92052     10.36055       2.50
```

continued on page 15

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
Fhealth	Fwealth	phi12	0.96065	0.03051	31.49

UN Poverty Data: STA431 S 2011 Final Exam 9
Latent regression model setting loadings to one

The 5 Endogenous Variables

Manifest	lifex	infmort	gnp1000	birthrate
Latent	Fhealth			

The 6 Exogenous Variables

Manifest					
Latent	Fwealth				
Error	epsilon	e1	e2	e3	e4

UN Poverty Data: STA431 S 2011 Final Exam 14
Latent regression model setting loadings to one

Optimization Start

Active Constraints	0	Objective Function	0.0176241672
Max Abs Gradient Element	0.0096430636	Radius	1

Iter	Restarts	Func Calls	Act Con	Objective Function	Obj Fun Change	Max Abs Gradient Element	Lambda	Actual Over Pred Change
1	0	2	0	0.01216	0.00547	0.00296	0	1.097
2	0	3	0	0.01207	0.000086	0.000199	0	1.019
3	0	4	0	0.01207	1.064E-6	0.000012	0	1.018
4	0	5	0	0.01207	1.596E-8	3.008E-6	0	1.017

continued on page 16

```
                         Optimization Results

Iterations                          4  Function Calls                    6
Jacobian Calls                      5  Active Constraints                0
Objective Function       0.0120694315  Max Abs Gradient Element  3.0080773E-6
Lambda                              0  Actual Over Pred Change   1.0166650766
Radius                   0.0010397183

ABSGCONV convergence criterion satisfied.

-------------------------------------------------------------------------------
                UN Poverty Data: STA431 S 2011 Final Exam                   15
             Latent regression model setting loadings to one

                          The CALIS Procedure
         Covariance Structure Analysis: Maximum Likelihood Estimation

         Fit Function                                         0.0121
         Goodness of Fit Index (GFI)                          0.9940
         GFI Adjusted for Degrees of Freedom (AGFI)           0.9404
         Root Mean Square Residual (RMR)                      2.5734
         Standardized Root Mean Square Residual (SRMR)        0.0067
         Parsimonious GFI (Mulaik, 1989)                      0.1657
         Chi-Square                                           1.0862
         Chi-Square DF                                             1
         Pr > Chi-Square                                      0.2973
         Independence Model Chi-Square                        399.83
         Independence Model Chi-Square DF                          6
-------------------------------------------------------------------------------
```

```
                UN Poverty Data: STA431 S 2011 Final Exam                   16
             Latent regression model setting loadings to one

                          The CALIS Procedure
         Covariance Structure Analysis: Maximum Likelihood Estimation

                  Manifest Variable Equations with Estimates


              lifex     =   1.0000 Fhealth   +  1.0000 e1
              infmort   =  -4.2928*Fhealth   +  1.0000 e2
              Std Err       0.1650 lambda2
              t Value     -26.0109
              gnp1000   =   1.0000 Fwealth   +  1.0000 e3
              birthrate =  -2.3190*Fwealth   +  1.0000 e4
              Std Err       0.2961 lambda4
              t Value      -7.8315

-------------------------------------------------------------------------------
```

continued on page 17

```
            UN Poverty Data: STA431 S 2011 Final Exam                 17
         Latent regression model setting loadings to one

                      The CALIS Procedure
  Covariance Structure Analysis: Maximum Likelihood Estimation

            Latent Variable Equations with Estimates


     Fhealth   =   1.8064*Fwealth   +  1.0000 epsilon
     Std Err       0.2317 gamma
     t Value       7.7958


                Variances of Exogenous Variables

                                                Standard
   Variable  Parameter      Estimate             Error    t Value

   Fwealth   phi            30.07656           8.40592       3.58
   epsilon   psi             8.20441           6.16622       1.33
   e1        omega1          1.44764           1.84498       0.78
   e2        omega2        184.16198          43.51832       4.23
   e3        omega3         35.43109           5.57604       6.35
   e4        omega4         25.92052          10.36055       2.50
```

Total Marks = 100 points