Name (Print): ______________________ ______________________

Last/Surname First /Given Name

Student Number: ______________________________________

Signature: ______________________________________

UNIVERSITY OF TORONTO MISSISSAUGA

April 2009 Examinations

STA431H5S

Professor Jerry Brunner

Duration: 3 hours

Aids allowed: Calculator (Any model is okay.)

You may be charged with an academic offence for possessing the following items during the writing of an exam unless otherwise specified: any unauthorized aids, including but not limited to calculators, cell phones, pagers, wristwatch calculators, personal digital assistants (PDAs), iPods, MP3 players, or any other device. If any of these items are in your possession in the area of your desk, please turn them off and put them with your belongings at the front of the room before the examination begins. A penalty may be imposed if any of these items are kept with you during the writing of your exam.

Please note, students are **not** allowed to petition to RE-WRITE a final examination.

Please write your answers in the examination booklet. Label the parts of your answers carefully!

15 points 1. The k-dimensional multivariate normal density is

$$f(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{k}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\Sigma}$ is symmetric and positive definite.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a random sample from a multivariate normal distribution with $\boldsymbol{\mu} = \mathbf{0}$. Write -2 times the log likelihood, and simplify until it is a function of $\boldsymbol{\Sigma}$ and $\mathbf{S} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'$ (and some constants). Show all your work. Because $\boldsymbol{\mu} = \mathbf{0}$, the symbol $\boldsymbol{\mu}$ should not appear in your answer. The sample mean $\overline{\mathbf{x}}$ also does not appear in the correct answer.

20 points 2. Consider the simple measurement model

$$\begin{aligned} X_1 &= F + e_1 \\ X_2 &= F + e_2, \end{aligned}$$

where X_1 and X_2 are observed variables, F is a latent variable that is independent of the measurement error terms e_1 and e_2, all the random variables are normally distributed with expected value zero, $Var(F) = \phi$, and

$$V\left[\begin{array}{c} e_1 \\ e_2 \end{array}\right] = \left[\begin{array}{cc} \omega_{1,1} & \omega_{1,2} \\ \omega_{1,2} & \omega_{2,2} \end{array}\right].$$

In *Model One*, we will assume that the error terms are independent; that is, $\omega_{1,2} = 0$.

(a) How do you know that the assumption $\omega_{1,2} = 0$ is necessary for the model to be identified? You do not need to do elaborate calculations; just cite a well-known rule.

(b) Assuming Model One, calculate the variance-covariance matrix of the observed variables. Express this matrix $\boldsymbol{\Sigma} = [\sigma_{i,j}]$ as a function of the model parameters. Show your work.

(c) Obtain explicit solutions of the identifying equations under Model One. Show your work. For later reference, you are writing $\boldsymbol{\theta} = \sigma^{-1}(\boldsymbol{\Sigma})$.

continued on page 3

(d) Recall the invariance principle of maximum likelihood estimation, which says that the MLE of a function of the parameter is that same function of the MLE. Denoting the maximum likelihood estimate of $\boldsymbol{\Sigma}$ by $\widehat{\boldsymbol{\Sigma}} = [\widehat{\sigma}_{i,j}]$, give an explicit formula for $\widehat{\omega}_{1,1}$. (We are still assuming Model One.) **Circle your answer.**

(e) Model One's assumption that $\omega_{1,2} = 0$ may well be unrealistic; it depends on how the data were collected. Under *Model Two*, $\omega_{1,2} > 0$. Of course Model Two is not identified, but it still may be correct. Calculate $\sigma_{1,2}$ as a function of the parameters of Model Two. Show your work and **Circle your answer.**

(f) By the Law of Large Numbers and continuous mapping, $\sigma^{-1}(\widehat{\boldsymbol{\Sigma}}) \rightarrow \sigma^{-1}(\boldsymbol{\Sigma})$, whether the model is right or not. Suppose that Model Two is correct, but you calculate the MLE under Model One (because it's really all you can do). What is the large-sample target of the variance estimate $\widehat{\omega}_{1,1}$?

(g) Under what condition will the variance estimate $\widehat{\omega}_{1,1}$ be negative for large samples? Your answer is an inequality in the parameters of Model Two.

(h) What is the lesson that this example is meant to illustrate? Give a short statement that shows you understand *why* you have been doing all these calculations.

15 points 3. The following is a multivariate regression model with measurement error in the independent variables but not the dependent variables. Let

$$\begin{aligned} \mathbf{X}_1 &= \boldsymbol{\xi} + \boldsymbol{\delta}_1 \\ \mathbf{X}_2 &= \boldsymbol{\xi} + \boldsymbol{\delta}_2 \\ \mathbf{Y} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \end{aligned}$$

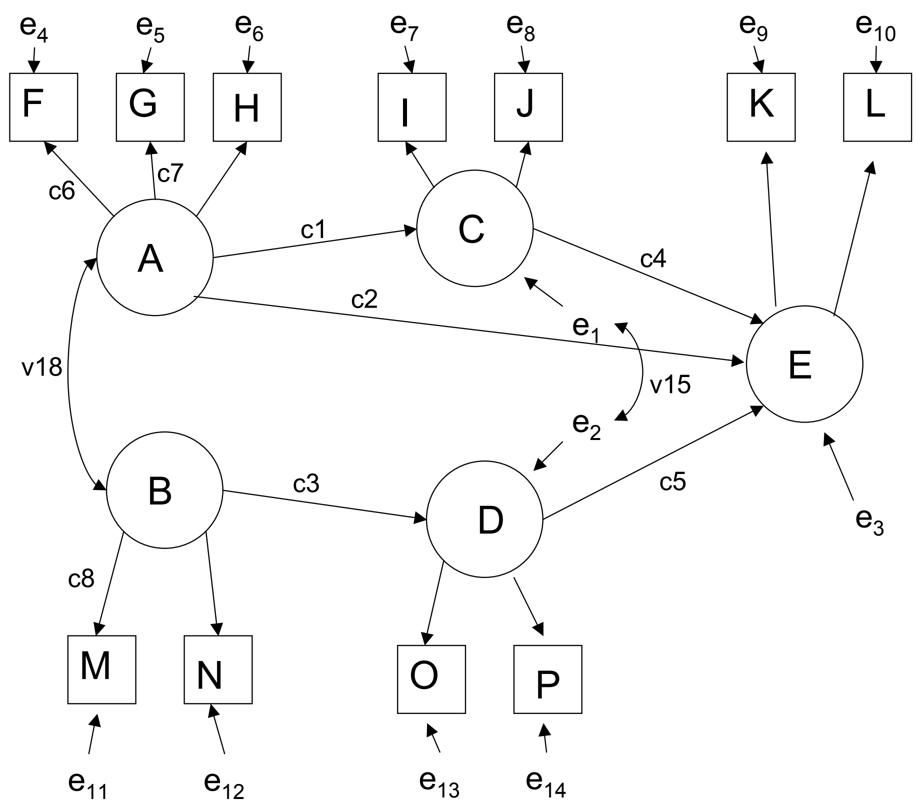
with $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ all independent, $V(\boldsymbol{\xi}) = \boldsymbol{\Phi}$, $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$, $V(\boldsymbol{\delta}_1) = \boldsymbol{\Omega}_1$, and $V(\boldsymbol{\delta}_2) = \boldsymbol{\Omega}_2$.

(a) Give the variance-covariance matrix of the observed variables $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{Y}$ in terms of the parameter matrices of the model.

(b) Is the model identified? Answer Yes or No and prove your answer **using matrix calculations**.

continued on page 4

10 points 4. In the path diagram below, $A \ldots P$ are random variables and $v_1, \ldots, v_{18}$ and $c_1, \ldots c_8$ are constants. Let $Var(e_j) = v_j$ for $j = 1, \ldots, 14$, $Cov(e_1, e_2) = v_{15}$, $Var(A) = v_{16}$, $Var(B) = v_{17}$ and $Cov(A, B) = v_{18}$. If a straight arrow is not marked with a coefficient, the coefficient equals one. All expected values are zero.



Can this model be identified from the covariance matrix? Answer Yes or No. If it is identified, briefly explain why, using a two-step argument. If it is not identified, state what change or changes are necessary to make it identified; then briefly explain why the modified model is identified, again using a two-step argument. Marks will be deducted for unnecessary changes. You don't need to present a formal proof or do any calculations.

continued on page 5

40 points

5. 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 6.

   (a) What is the parameter $\boldsymbol{\theta}$ for this model? **Give your answer in the form of a list of names from the SAS job.**

   (b) Make a path diagram of the model. If an arrow does not have a coefficient on it, the coefficient equals one. If an arrow is missing, the coefficient equals zero. These rules apply to curved double-headed arrows as well as to straight arrows. **Use the variable names and parameter names from the SAS job**.

   (c) Does this 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.

   (d) What are the maximum likelihood estimates of γ_1, γ_2, γ_3 and γ_4? Give four numbers from the printout.

   (e) Why does the MLE of γ_4 make sense? Answer in one simple sentence.

   (f) Prove that the model is identified. Show your work. In order to proceed, you will have to make assumptions about the value(s) of certain parameter(s). **When you make such an assumption, state it and underline the statement.**

   (g) Clearly, the model is not identified at all points in the parameter space. Give an example of two values of $\boldsymbol{\theta}$ that yield the same covariance matrix $\boldsymbol{\Sigma}$.

   (h) In your view, is the true parameter value likely to be at one of the points where the model is not identified? Answer Yes or No and briefly explain. Does the printout support your position?

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

Total Marks = 100 points

continued on page 6

```
/********************** finalpoverty2.sas ******************/
options linesize=79 noovp formdlim='-';
title 'UN Poverty Data: Final Exam Regression';

data misery;
     infile 'poverty.data';
     input birthrat deathrat infmort lifexM lifexF gnp;
     gnp1000 = gnp/1000; /* In thousands of dollars */

proc calis cov;
     var gnp1000 lifexM lifexF birthrat infmort;
     lineqs
          gnp1000 = Fgnp + delta,
          lifexM  = gamma1 Fgnp + e1,
          lifexF  = gamma2 Fgnp + e2,
          birthrat = gamma3 Fgnp + e3,
          infmort  = gamma4 Fgnp + e4;
     std
          Fgnp = phi,
          delta = omega,
          e1-e4 = 4 * psi: ;
     cov e1 e2 e4 = 3 * PSIij: ;
     bounds 0.0 < phi omega psi1-psi4;
```

continued on page 7

Parts of finalpoverty2.lst

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

                    UN Poverty Data: Final Exam Regression
                          The 5 Endogenous Variables

    Manifest        gnp1000   lifexM    lifexF    birthrat  infmort
    Latent


                          The 6 Exogenous Variables

    Manifest
    Latent          Fgnp
    Error           e1        e2        e3        e4        delta


                              The CALIS Procedure
         Covariance Structure Analysis: Maximum Likelihood Estimation

            Observations          91    Model Terms              1
            Variables              5    Model Matrices           4
            Informations          15    Parameters              13

                             Optimization Results

Iterations                           5  Function Calls                       10
Jacobian Calls                       6  Active Constraints                    0
Objective Function        0.0157591458  Max Abs Gradient Element   2.136072E-6
Lambda                               0  Actual Over Pred Change   1.1375548744
Radius                    0.0026793849

ABSGCONV convergence criterion satisfied.
```

continued on page 8

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

             UN Poverty Data: Final Exam Regression                          6

                        The CALIS Procedure
         Covariance Structure Analysis: Maximum Likelihood Estimation

         Fit Function                                                0.0158
         Goodness of Fit Index (GFI)                                 0.9938
         GFI Adjusted for Degrees of Freedom (AGFI)                  0.9534
         Root Mean Square Residual (RMR)                             1.5585
         Parsimonious GFI (Mulaik, 1989)                             0.1988
         Chi-Square                                                  1.4183
         Chi-Square DF                                                    2
         Pr > Chi-Square                                             0.4921
         Independence Model Chi-Square                               716.00
         Independence Model Chi-Square DF                                10
         RMSEA Estimate                                              0.0000
         RMSEA 90% Lower Confidence Limit                                 .
         RMSEA 90% Upper Confidence Limit                            0.1888
         ECVI Estimate                                               0.3253
         ECVI 90% Lower Confidence Limit                                  .
         ECVI 90% Upper Confidence Limit                             0.4067
         Probability of Close Fit                                    0.5651
         Bentler's Comparative Fit Index                             1.0000
         Normal Theory Reweighted LS Chi-Square                      1.4072
         Akaike's Information Criterion                             -2.5817
         Bozdogan's (1987) CAIC                                     -9.6034
         Schwarz's Bayesian Criterion                               -7.6034
         McDonald's (1989) Centrality                                1.0032
         Bentler & Bonett's (1980) Non-normed Index                  1.0041
         Bentler & Bonett's (1980) NFI                               0.9980
         James, Mulaik, & Brett (1982) Parsimonious NFI              0.1996
         Z-Test of Wilson & Hilferty (1931)                          0.0086
         Bollen (1986) Normed Index Rho1                             0.9901
         Bollen (1988) Non-normed Index Delta2                       1.0008
         Hoelter's (1983) Critical N                                    382
```

continued on page 9

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

                   UN Poverty Data: Final Exam Regression                      7

                             The CALIS Procedure
        Covariance Structure Analysis: Maximum Likelihood Estimation

                  Manifest Variable Equations with Estimates


                gnp1000  =   1.0000 Fgnp     +  1.0000 delta
                lifexM   =   1.6600*Fgnp     +  1.0000 e1
                Std Err      0.2182 gamma1
                t Value      7.6092
                lifexF   =   1.9568*Fgnp     +  1.0000 e2
                Std Err      0.2508 gamma2
                t Value      7.8007
                birthrat =  -2.3435*Fgnp     +  1.0000 e3
                Std Err      0.2987 gamma3
                t Value     -7.8468
                infmort  =  -7.7661*Fgnp     +  1.0000 e4
                Std Err      1.0356 gamma4
                t Value     -7.4992


                     Variances of Exogenous Variables

                                                     Standard
           Variable Parameter        Estimate           Error    t Value

           Fgnp     phi              29.76160         8.35607       3.56
           e1       psi1             12.61527         5.29287       2.38
           e2       psi2              9.93937         6.66231       1.49
           e3       psi3             24.20877         9.96144       2.43
           e4       psi4            348.92294       123.27590       2.83
           delta    omega            35.74606         5.59044       6.39
```

continued on page 10

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
e1	e2	PSIij1	9.64976	5.83516	1.65
e1	e4	PSIij2	-37.56349	24.18587	-1.55
e2	e4	PSIij3	-39.52078	27.43491	-1.44

Total Marks = 100 points