NAME (PRINT): ______________________________________
Last/Surname First /Given Name

STUDENT #: ____________________ **SIGNATURE:** ____________________

UNIVERSITY OF TORONTO MISSISSAUGA
APRIL 2015 FINAL EXAMINATION
STA431H5S
Structural Equation Models
Jerry Brunner
Duration - 3 hours
Aids: Calculator Model(s): Any calculator is okay; Formula sheet will be 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. Clear, sealable, plastic bags have been provided for all electronic devices with storage, including but not limited to: cell phones, tablets, laptops, calculators, and MP3 players. Please turn off all devices, seal them in the bag provided, and place the bag under your desk for the duration of the examination. You will not be able to touch the bag or its contents until the exam is over.

If, during an exam, any of these items are found on your person or in the area of your desk other than in the clear, sealable, plastic bag; you may be charged with an academic offence. A typical penalty for an academic offence 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	12	
2	14	
3	10	
4	13	
5	13	
6	10	
7	8	
8	8	
9	12	
Total = 100 Points		

12 points

1. This simple example illustrates the reason for the entire course. Independently for $i = 1, \ldots, n$, let

$$\begin{aligned} Y_i &= \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon_i \\ W_i &= X_{i,1} + e_i, \end{aligned}$$

where $V\begin{pmatrix} X_{i,1} \\ X_{i,2} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}$, $V(\epsilon_i) = \psi$, $V(e_1) = \omega$, all the expected values are zero, and the error terms ϵ_i and e_i are independent of one another, and also independent of $X_{i,1}$ and $X_{i,2}$. The variable $X_{i,1}$ is latent, while the variables W_i, Y_i and $X_{i,2}$ are observable. What people usually do in situations like this is fit a model like $Y_i = \beta_1 W_i + \beta_2 X_{i,2} + \epsilon_i$, and test $H_0 : \beta_2 = 0$. That is, they ignore the measurement error in variables for which they are "controlling."

(a) Suppose $H_0 : \beta_2 = 0$ is true. Does the ordinary least squares estimator

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^n W_i^2 \sum_{i=1}^n X_{i,2} Y_i - \sum_{i=1}^n W_i X_{i,2} \sum_{i=1}^n W_i Y_i}{\sum_{i=1}^n W_i^2 \sum_{i=1}^n X_{i,2}^2 - (\sum_{i=1}^n W_i X_{i,2})^2}$$

converge to the true value of $\beta_2 = 0$ as $n \to \infty$ everywhere in the parameter space? Answer Yes or No and show your work. (And don't forget to answer the question at the bottom of the next page).

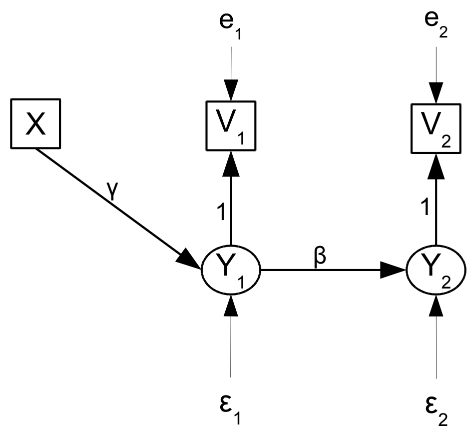
continued on page 3

(b) Under what conditions (that is, for what values of other parameters) does $\widehat{\beta}_2 \stackrel{a.s.}{\to} 0$ when $\beta_2 = 0$?

continued on page 4

14 points

2. In a study of maternal behaviour in rats, mother rats were injected with estrogen (a sex hormone), and the amount of estrogen in their blood was measured by a blood sample. Then their maternal behaviour (nursing, licking, retrieving their young) was recorded. In the path diagram below, the symbol on the arrow from X to Y_1 looks like the letter Y, but actually it's the Greek letter γ (gamma).



X is drug dose.

Y_1 is true amount of drug in the animal's blood stream.

V_1 is measured amount of drug in the animal's blood stream.

Y_2 is the animal's maternal impulse, or tendency to care for her young.

V_2 is the animal's observed maternal behaviour.

(a) What is the parameter vector $\boldsymbol{\theta}$ for this model? All the variables are centered, so there are no expected values or intercepts.

(b) Does this problem pass the test of the Parameter Count Rule? Answer Yes or No and give the numbers.

(c) Why is it reasonable to assume $\gamma > 0$?

(d) The point of the study is β. Show that this parameter identifiable provided $\gamma > 0$. *Don't do more work than you have to.*

continued on page 5

continued on page 6

10 points

3. Independently for $i = 1, \ldots, n$, let

$$\begin{array}{rcl} D_{i,1} & = & F_{i,1} + e_{i,1} \\ D_{i,2} & = & F_{i,2} + e_{i,2} \end{array} \quad V\left(\begin{array}{c} F_{i,1} \\ F_{i,2} \end{array} \right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{array} \right) \quad V\left(\begin{array}{c} e_{i,1} \\ e_{i,2} \end{array} \right) = \left(\begin{array}{cc} \omega_1 & 0 \\ 0 & \omega_2 \end{array} \right)$$

Suppose the reliability of $D_{i,1}$ as a measure of $F_{i,1}$ is r_1, and the reliability of $D_{i,2}$ as a measure of $F_{i,2}$ is r_2. If $Corr(D_{i,1}, D_{i,2}) = \rho$, what is $Corr(F_{i,1}, F_{i,2})$? Show your work and **circle your final answer**. The answer is an expression in just ρ, r_1 and r_2. It's a way of correcting correlations for measurement error if you know the reliabilities.

continued on page 7

13 points

4. In this special case of the latent variable model, there are no exogenous variables. Independently for $i = 1, \ldots, n$, let $\mathbf{Y}_i = \boldsymbol{\beta}\mathbf{Y}_i + \boldsymbol{\epsilon}_i$, where $\mathbf{Y}_i$ and $\boldsymbol{\epsilon}_i$ are $q \times 1$ random vectors with $V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}$. The $q \times q$ covariance matrix $\boldsymbol{\Psi}$ is strictly positive definite.

   (a) Calculate $V(\mathbf{Y}_i)$. Show your work.

   (b) Prove that the existence of $(\mathbf{I} - \boldsymbol{\beta})^{-1}$ is implied by the model.

continued on page 8

13 points

5. The general *unrestricted* exploratory factor analysis model can be written $\mathbf{D}_i = \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i$, where $\boldsymbol{\Lambda}$ is a $k \times p$ matrix of constants, $V(\mathbf{F}_i) = \boldsymbol{\Phi}$, $V(\mathbf{e}_i) = \boldsymbol{\Omega}$, and the random vectors $\mathbf{F}_i$ and $\mathbf{e}_i$ are independent. As usual, the covariance matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ are positive definite.

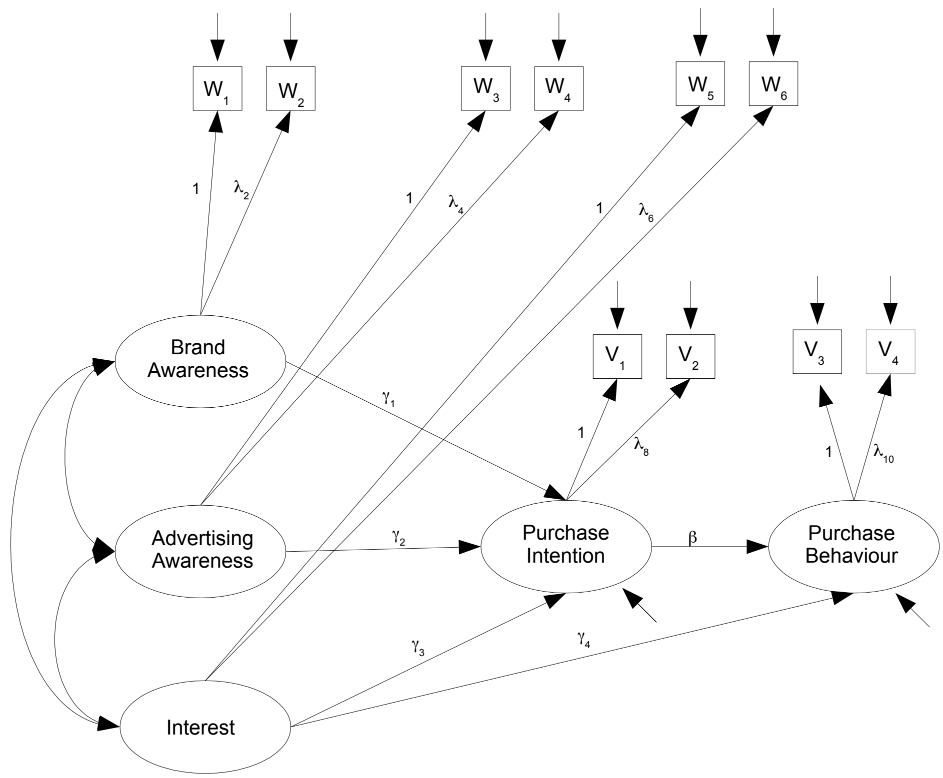
   (a) Write down $\boldsymbol{\Sigma} = V(\mathbf{D}_i)$. It's so quick that you need not show any work.

   (b) The problem is that $\boldsymbol{\Phi}$ could be absolutely anything, and you can't tell from $\boldsymbol{\Sigma}$. Ignoring redundant elements of the covariance matrices, the parameter vector could be written $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Omega})$. Let $\mathbf{Q}$ be an arbitrary $p \times p$ symmetric positive definite matrix. There is another parameter vector $\boldsymbol{\theta}_2 = (\boldsymbol{\Lambda}_2, \mathbf{Q}, \boldsymbol{\Omega})$ yielding exactly the same $\boldsymbol{\Sigma}$. Give the matrix $\boldsymbol{\Lambda}_2$. Show your work. **Circle your answer**.

continued on page 9

10 points

6. Compare this path diagram with the general structural equation model on the formula sheet.



Give the matrix $\mathbf{\Gamma}$. Your answer is a matrix in which each element is either a zero or a symbol from the path diagram.

continued on page 10

8 points

7. Please look again at the path diagram of Question 6. Briefly explain why the parameters of this model are identifiable. Cite specific rules *by letter and number*. Assume that if an arrow is shown, the coefficient is not zero. You have a lot more room than you need for a good answer. You do *not* have to explain in detail why a particular rule applies.

continued on page 11

8 points

8. In a reaction time study, subjects are seated at a screen. A light flashes on the screen, and they press a key as fast as they can; the time between the light flash and the key press is recorded automatically.

   After some warmup trials, the subjects do the task 50 times, so 50 reaction times are recorded. The 50 times are divided randomly into two sets of 25, and then the median is calculated for each set. In the end, each subject produces two median reaction times.

   The scientists locate sample of university student volunteers whose parents are also available to do the experiment. When all the observable data have been collected, there is a data file with n lines of data, one for each student. Each line of data has 6 numbers. There are two median reaction times for the student, two for the student's mother and two for the student's father.

   Make a path diagram. Here are some guidelines.

   - Include the error terms. You should assume that the errors are all independent, and independent of the exogenous variables.
   - Write coefficients on the arrows. If you leave a coefficient off, it means the coefficient equals one.
   - This is an *original* model, not re-parameterized.
   - Your model should represent reasonable ideas about heredity, but my answer is quite unsophisticated so don't worry.
   - There is definitely more than one right answer, but all else being equal, simpler models are better.
   - You may write some comments under the path diagram if you wish, but no comments are necessary for full marks.

continued on page 12

12 points

9. Recall the study of farm co-op managers, in which several latent variables are measured twice. The response variable is job performance.

```
     label w11 = 'Knowledge 1'
           w12 = 'Knowledge 2'
           w21 = 'Profit-Loss Orientation 1'
           w22 = 'Profit-Loss Orientation 2'
           w31 = 'Job Satisfaction 1'
           w32 = 'Job Satisfaction 2'
           edu = 'Formal education' /* Assumed measured without error */
            v1 = 'Job Performance 1'
            v2 = 'Job Performance 2';

proc calis pshort nostand pcorr vardef=n;
     var w11 w12 w21 w22 w31 w32 edu v1 v2;
     lineqs
          Fperform = beta1 Fknowledge + beta2 Fprofitloss + beta3 Fsatisf
                     + beta4 edu + epsilon,
          w11 = Fknowledge + e1,
          w12 = Fknowledge + e2,
          w21 = Fprofitloss + e3,
          w22 = Fprofitloss + e4,
          w31 = Fsatisf + e5,
          w32 = Fsatisf + e6,
          v1  = Fperform + e7,
          v2  = Fperform + e8;
     variance
          Fknowledge=phi11, Fprofitloss=phi22, Fsatisf=phi33, edu=phi44,
          epsilon = psi, e1-e8 = 8 * omega__ ;
     cov
          Fknowledge Fprofitloss = phi12, Fknowledge Fsatisf = phi13,
          Fknowledge edu = phi14,
          Fprofitloss Fsatisf = phi23, Fprofitloss edu = phi24,
          Fsatisf edu = phi34;
     bounds 0.0 < phi11 phi22 phi33 phi44 psi omega1-omega6;
     simtests AllVars = [b1 b2 b3 b4];
     b1 = beta1 ; b2 = beta2 ; b3 = beta3; b4 = beta4;
     simtests ReliabilityDifference = [d]; d = omega1-omega2;
```

The next several pages contain partial output, followed by the questions.

continued on page 13

431s15A8: Re-constructed farm co-op manager data
Using Warren White and Fuller 1974, and Joreskog 1978

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Summary		
Modeling Info	Number of Observations	98
	Number of Variables	9
	Number of Moments	45
	Number of Parameters	23
	Number of Active Constraints	0
	Baseline Model Function Value	2.8746
	Baseline Model Chi-Square	281.7096
	Baseline Model Chi-Square DF	36
	Pr > Baseline Model Chi-Square	<.0001
Absolute Index	Fit Function	0.2774
	Chi-Square	27.1819
	Chi-Square DF	22
	Pr > Chi-Square	0.2044
	Z-Test of Wilson & Hilferty	0.8273
	Hoelter Critical N	122
	Root Mean Square Residual (RMR)	0.0065
	Standardized RMR (SRMR)	0.0638
	Goodness of Fit Index (GFI)	0.9455
Parsimony Index	Adjusted GFI (AGFI)	0.8885
	Parsimonious GFI	0.5778
	RMSEA Estimate	0.0490
	RMSEA Lower 90% Confidence Limit	0.0000
	RMSEA Upper 90% Confidence Limit	0.1024
	Probability of Close Fit	0.4719
	ECVI Estimate	0.8001
	ECVI Lower 90% Confidence Limit	0.7727
	ECVI Upper 90% Confidence Limit	0.9855
	Akaike Information Criterion	73.1819
	Bozdogan CAIC	155.6362
	Schwarz Bayesian Criterion	132.6362
	McDonald Centrality	0.9739
Incremental Index	Bentler Comparative Fit Index	0.9789
	Bentler-Bonett NFI	0.9035
	Bentler-Bonett Non-normed Index	0.9655
	Bollen Normed Index Rho1	0.8421

continued on page 14

431s15A8: Re-constructed farm co-op manager data
Using Warren White and Fuller 1974, and Joreskog 1978

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Linear Equations

Fperform	=	0.3468 (**)	Fknowledge	+ 0.1712 (**)	Fprofitloss	+ 0.0498 (ns)	Fsatisf + 0.0726 (ns) edu	+ 1.0000	epsilon
w11	=	1.0000	Fknowledge	+ 1.0000	e1				
w12	=	1.0000	Fknowledge	+ 1.0000	e2				
w21	=	1.0000	Fprofitloss	+ 1.0000	e3				
w22	=	1.0000	Fprofitloss	+ 1.0000	e4				
w31	=	1.0000	Fsatisf	+ 1.0000	e5				
w32	=	1.0000	Fsatisf	+ 1.0000	e6				
v1	=	1.0000	Fperform	+ 1.0000	e7				
v2	=	1.0000	Fperform	+ 1.0000	e8				

Effects in Linear Equations

Variable	Predictor	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Fperform	Fknowledge	beta1	0.34683	0.13166	2.6342	0.0084
Fperform	Fprofitloss	beta2	0.17122	0.07909	2.1649	0.0304
Fperform	Fsatisf	beta3	0.04984	0.05040	0.9889	0.3227
Fperform	edu	beta4	0.07260	0.04420	1.6427	0.1004
w11	Fknowledge		1.00000			
w12	Fknowledge		1.00000			
w21	Fprofitloss		1.00000			
w22	Fprofitloss		1.00000			
w31	Fsatisf		1.00000			
w32	Fsatisf		1.00000			
v1	Fperform		1.00000			
v2	Fperform		1.00000			

Estimates for Variances of Exogenous Variables

Variable Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Latent	**Fknowledge**	phi11	0.03017	0.00761	3.9646	<.0001
	Fprofitloss	phi22	0.07593	0.01814	4.1859	<.0001
	Fsatisf	phi33	0.07158	0.01299	5.5088	<.0001
Observed	**edu**	phi44	0.09363	0.01338	7.0000	<.0001
Disturbance	**epsilon**	psi	0.00550	0.00193	2.8468	0.0044
Error	**e1**	omega1	0.05026	0.00895	5.6128	<.0001
	e2	omega2	0.02996	0.00685	4.3725	<.0001
	e3	omega3	0.09132	0.01802	5.0671	<.0001
	e4	omega4	0.08248	0.01713	4.8148	<.0001
	e5	omega5	0.04147	0.00985	4.2117	<.0001

continued on page 15

431s15A8: Re-constructed farm co-op manager data
Using Warren White and Fuller 1974, and Joreskog 1978

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Estimates for Variances of Exogenous Variables

Variable Type	Variable	Parameter	Estimate	Standard Error	t Value	Pr > |t|
	e6	omega6	0.02952	0.00892	3.3082	0.0009
	e7	omega7	0.00878	0.00182	4.8353	<.0001
	e8	omega8	0.00596	0.00157	3.8086	0.0001

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value	Pr > |t|
Fknowledge	**Fprofitloss**	phi12	0.02564	0.00814	3.1491	0.0016
Fknowledge	**Fsatisf**	phi13	0.00546	0.00668	0.8167	0.4141
Fknowledge	**edu**	phi14	0.01875	0.00710	2.6423	0.0082
Fprofitloss	**Fsatisf**	phi23	-0.00623	0.01042	-0.5978	0.5500
Fprofitloss	**edu**	phi24	0.03388	0.01121	3.0223	0.0025
Fsatisf	**edu**	phi34	-0.00687	0.00924	-0.7438	0.4570

Simultaneous Tests

Simultaneous Test	Parametric Function	Function Value	DF	Chi-Square	p Value
AllVars			4	63.27343	<.0001
	b1	0.34683	1	6.93896	0.0084
	b2	0.17122	1	4.68658	0.0304
	b3	0.04984	1	0.97791	0.3227
	b4	0.07260	1	2.69851	0.1004
ReliabilityDifference			1	3.35235	0.0671
	d	0.02030	1	3.35235	0.0671

continued on page 16

(a) Based strictly on the $\alpha = 0.05$ significance level, does the model appear to fit the data adequately?

i. Answer the question Yes or No.

ii. Give the value of a test statistic and a p-value that supports your conclusion. These are two numbers from the printout.

(b) Controlling for formal education, knowledge and profit-loss orientation, is there evidence that job satisfaction is related to job performance? Base your answers strictly on the $\alpha = 0.05$ significance level.

i. Answer the question Yes or No.

ii. Give the value of a test statistic and a p-value that supports your conclusion. These are two numbers from the printout.

iii. If the answer is Yes (and *only* if the answer is Yes), say whether satisfaction is positively related to performance, or negatively related.

Total Marks = 100 points