

BUEC 333: Study Questions for the Final Exam

1. All study questions for the midterm
  2. All assignments
  3. Studenmund Ch1, Questions: 4, 5, 6, 7, 8, 9, 10, 11, 12, 13
  4. Studenmund Ch2, Questions: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13
  5. Studenmund Ch3, Questions: 1, 2, 3, 5, 8, 9, 10
  6. Studenmund Ch4, Questions: 1, 2, 3, 4, 5, 7, 9, 10, 11
  7. Studenmund Ch5, Questions: 1, 2, 3, 6, 8, 9, 10, 11, 13, 14
  8. Studenmund Ch6, Questions: 1, 2, 3, 4, 6, 11, 12, 13, 15, 16
  9. Studenmund Ch7, Questions: 1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16a-f
  10. Studenmund Ch8, Questions: 1ab, 2, 3, 5, 6, 7, 8, 9, 10, 11, 13, 15
  11. Studenmund Ch9, Questions: 1, 2, 5abce, 7, 8, 9, 11, 13, 14a-d, 15
  12. Studenmund Ch10, Questions: 1, 2, 4, 9, 10, 11, 12, 14, 15
13. Suppose that on any given trading day, a stock's price can increase, decrease, or remain unchanged. Likewise, on any given day, the weather can be sunny or cloudy. You have the following information about the joint distribution of weather and stock price changes:

	Price Decreases	Price Unchanged	Price Increases
Sunny Weather	3/20	3/10	<b>3/20</b>
Cloudy Weather	1/10	1/5	1/10

- a. What is the joint probability of sunny weather and a price increase?
  - i. probabilities sum to 1, so that probability is 1 minus all the others (remember that  $1/5=2/10$ ,  $3/20=1.5/10$ ):  $1.5/10$ .
- b. If the average price increase is \$1 and the average price decrease is \$1, what is the expected price change on any given day?
  - i. this requires use of the marginal probabilities, which do NOT condition on the weather. marginals are  $2.5/10$ ,  $5/10$ ,  $2.5/10$  for decrease, unchanged and increased.
  - ii. the expected value is the probability weighted average:  $2.5/10*(-1) + 5/10*0 + 2.5/10*1=0$
- c. It is sunny today. What is the probability of a price increase?
  - i. the conditional probability is equal to the joint probability divided by the marginal probability:  $P[\text{price increase} | \text{sunny}] = P[\text{price increase \& sunny}] / P[\text{sunny}] = (1.5/10) / (6/10) = 1.5/6$
- d. Explain the difference between your answer to part a and your answer to part c.
  - i. a was about the joint probability; c was about a conditional probability
- e. What can you say about the relationship between weather and stock price changes? Explain.

- i. the probability of sunny weather is 3/5 no matter the price change. they are independent.

14. Suppose  $Y_i = \alpha + \beta X_i + \varepsilon_i$  and that  $\varepsilon_i = u_i + v_i + kX_i$

- a. Assume  $k=0$ , and that  $u$  and  $v$  are each distributed as *standard normals*. What is the distribution of  $\varepsilon_i$ ? Would an OLS regression of  $Y$  on  $X$  yield unbiased estimates of the marginal effect of  $X$  on  $Y$ ? Would the estimated coefficient be efficient?
  - i. epsilon is the sum of 2 standard normals. since  $u$  and  $v$  are both standard normals, they are independent of each other. using the rules for means and variances, the mean of epsilon is 0 and the variance is 2. since they are normals, we have  $\varepsilon_i \sim N(0,2)$
  - ii. This error term has a mean of 0 no matter what  $X$  is, and has the same variance (2) for all observations, and no dependence across observations. OLS is unbiased and efficient.
- b. Assume  $k=0$ , and that  $u$  is distributed as a normal with mean 0 and variance  $2X$ , and that  $v$  is distributed as a *standard normals*. Would an OLS regression of  $Y$  on  $X$  yield unbiased estimates of the marginal effect of  $X$  on  $Y$ ? Would the estimated coefficient be efficient?
  - i. the interpretation I intended is that  $X$  is the same variable as in the equation. The right notation would have been "is distributed as a normal with mean 0 and variance  $2X_i$ ". Then, using the rules for means and variances, and the fact that I am adding normals, I have  $\varepsilon_i \sim N(0,1+2X_i)$
  - ii. This error is mean 0 no matter what  $X$  is, but its variance is different for different  $X$  values. OLS is unbiased but not efficient.
- c. Assume  $k=1$ , and that  $u$  and  $v$  are each distributed as *standard normals*. Would an OLS regression of  $Y$  on  $X$  yield unbiased estimates of the marginal effect of  $X$  on  $Y$ ? Would the estimated coefficient be efficient?
  - i.  $\varepsilon_i = X_i + u_i + v_i = X_i + N(0,1) + N(0,1) = X_i + N(0,2)$
  - ii. This error term covaries with  $X$ . The covariance is given by
  - iii. 
$$\begin{aligned} \text{Cov}(X_i, X_i + N(0,2)) &= E[(X_i - E[X])(X_i + N(0,2) - E[X + N(0,2)])] \\ &= E[(X_i - E[X])(X_i - E[X] + N(0,2))] \\ &= V(X) + E[(X_i - E[X])(N(0,2))] \\ &= V(X) + 0 \end{aligned}$$

The covariance is not zero, so OLS is biased. In addition, the variance of the errors changes with  $X$ , so the OLS estimator is inefficient.

- d. Assume  $k=0$ , and that  $u$  is distributed as a *standard normal*, and that  $v$  is distributed as a normal with mean  $X$  and variance 1. Would an OLS regression of  $Y$  on  $X$  yield unbiased estimates of the marginal effect of  $X$  on  $Y$ ? Would the estimated coefficient be efficient?
  - i.  $\varepsilon_i = u_i + v_i = N(0,1) + N(X_i,1) = N(X_i,2)$
  - ii. this model is identical to that in c, so the covariance is identical to that in c, and OLS is both biased and inefficient.
- e. Assume that  $k$ ,  $u$  and  $v$  are all distributed as a coin-flip with 50% probability of being -1 and 50% probability of being 1. Would an OLS regression of  $Y$  on  $X$  yield

unbiased estimates of the marginal effect of  $X$  on  $Y$ ? Would the estimated coefficient be efficient?

i. Make a table of the values of the error term:

	K=-1				K=1			
	U=-1		U=1		U=-1		U=1	
	V=1	V=-1	V=1	V=-1	V=1	V=-1	V=1	V=-1
epsilon	-X	-2-X	2-X	-X	X	-2+X	2+X	X

Each outcome has an equal probability. The expected value is thus 0 (probability weighted average of the values of epsilon) no matter what value  $X$  is. Thus, OLS is unbiased. However, the variance of epsilon does depend on  $X$ : it is bigger when  $X$  is far from zero. So, OLS is not efficient.

15. Suppose  $Y_i = \alpha + \beta X_i + \varepsilon_i$  and that  $\varepsilon_i = u_i + 0.25 * (\varepsilon_{i-2} + \varepsilon_{i-1} + \varepsilon_{i+1} + \varepsilon_{i+2})$  and that  $u$  is distributed as a standard normal. Would an OLS regression of  $Y$  on  $X$  yield unbiased estimates of the marginal effect of  $X$  on  $Y$ ? Would the estimated coefficient be efficient? Would the OLS variance of the estimated coefficient be correct? If not, how could you obtain the correct variance? Could you improve on this estimator to get a lower variance estimate of the coefficient?

- Each epsilon has an expectation of zero, no matter the value of  $X$ . So, OLS is unbiased.
- Each epsilon depends on 4 other epsilons. therefore the covariance between epsilons is not zero for some of them. Therefore, OLS is not efficient.
- Because OLS coefficient variances are calculated assuming that epsilons do not covary with each other, the OLS coefficient variances are wrong.
- The question of obtaining the correct variance is too hard for you guys. In fact, you'd need to allow for BOTH heteroskedasticity (White variances) and correlation across observations (Newey-West).

16. Suppose  $Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \varepsilon_i$  and  $X_i = \alpha_1 Z_i + k u_i$  and  $W_i = \alpha_2 Z_i + k v_i$  and that  $Z, u, v$  and  $\varepsilon_i$  are distributed as standard normals. What is the covariance of  $X$  and  $W$ ?

$$\begin{aligned}
 \text{a. } \text{Cov}(X_i, W_i) &= E\left[(\alpha_1 Z_i + k u_i - E[\alpha_1 Z_i + k u_i])(\alpha_2 Z_i + k v_i - E[\alpha_2 Z_i + k v_i])\right] \\
 &= E\left[(\alpha_1 Z_i + k u_i - E[\alpha_1 Z_i])(\alpha_2 Z_i + k v_i - E[\alpha_2 Z_i])\right] \\
 &= E\left[(\alpha_1 Z_i - E[\alpha_1 Z_i])(\alpha_2 Z_i - E[\alpha_2 Z_i])\right] \\
 &= \text{Cov}(\alpha_1 Z_i, \alpha_2 Z_i) = \alpha_1 \alpha_2 \text{Var}(Z_i) = \alpha_1 \alpha_2
 \end{aligned}$$

Note that the third equality follows from the fact that  $u$  and  $v$  are standard normals, and the fourth equality uses the definition of a covariance and the fact that  $Z$  is a standard normal.

- What values of the parameters result in a perfect multicollinearity problem.
  - $X$  and  $W$  covary. But they are only linear in each other if  $k=0$ . In this case,

$$X_i = \alpha_1 Z_i, W_i = \alpha_2 Z_i, \text{ so } X_i = \frac{\alpha_1}{\alpha_2} W_i$$

- What values of the parameters would result in "some" multicollinearity that might frustrate a regression of  $Y$  on  $X$  and  $W$ ?

- i. if  $k$  were 'small' relative to  $\alpha_1$  and  $\alpha_2$ , then there is some multicollinearity.
  - d. Suppose that the covariance of  $u$  and  $\varepsilon_i$  were nonzero (in this case, they would not be standard normals, they would be correlated normals). Would this cause a problem?
    - i. if  $u$  and  $\varepsilon$  were correlated, then  $X$  and  $\varepsilon$  would be correlated (because  $X$  is linear in  $u$ ). That is, the expectation of  $\varepsilon$  would depend on  $X$ , which would imply that OLS would be biased.
17. Consider the following cross-tab of age and Employment Equity Group for Canadian-born people aged 25-64 with nonmissing wages > 100 in Vancouver (AGEGRP is coded as in the data codebook; EEGROUP is 1=white 2=vismin 3=aboriginal) :

Count	AGEGRP								Total
	9	10	11	12	13	14	15	16	
EEGROU 1	1526	1557	1600	1670	1774	1554	1213	622	11516
EEGROU 2	463	274	201	139	116	49	33	20	1295
EEGROU 3	67	70	51	68	55	43	31	20	405
Total	2056	1901	1852	1877	1945	1646	1277	662	13216

Does knowing EEGROUP tell you anything about age?

*Compute the table of conditional probabilities for age groups, conditioning on EEGROUP. It shows that Aboriginals are younger than non-Aboriginals---the conditional probability of being in a younger age group is higher for them than for other groups.*

*These are the conditional probabilities*

	9	10	11	12	13	14	15	16
1	0.13	0.14	0.14	0.15	0.15	0.13	0.11	0.05
2	0.36	0.21	0.16	0.11	0.09	0.04	0.03	0.02
3	0.17	0.17	0.13	0.17	0.14	0.11	0.08	0.05

Are age and EEGROUP correlated? What does this mean? Define a dummy variable called Aboriginal equal to 1 if EEGROUP=3.

*Correlation has meaning when the values of the variables have meaning. The values of AGEGRP do have some meaning: they are higher when age is higher. The values of EEGROUP (1,2,3) don't have meaning. But, the value 1 for the Aboriginal dummy does have mean: such people have 1 unit of Aboriginalness.*

Do age and Aboriginal have nonzero covariance?

*Can compute the conditional probabilities, given Aboriginal status*

	9	10	11	12	13	14	15	16
Aboriginal 0	0.16	0.14	0.14	0.14	0.15	0.13	0.10	0.05
Aboriginal 1	0.17	0.17	0.13	0.17	0.14	0.11	0.08	0.05

This shows that Aboriginals are overrepresented in younger age groups.

*To compute the covariance, first make a cross-tab of Aboriginal and AGEGRP*

cross-tab	9	10	11	12	13	14	15	16	
Aboriginal 0	1989	1831	1801	1809	1890	1603	1246	642	12811

Then, convert the cross-tab into probabilities of being in each category by dividing by the total number of people:

probabilities		9	10	11	12	13	14	15	16
Aboriginal	0	0.1505	0.1385	0.1363	0.1369	0.1430	0.1213	0.0943	0.0486
	1	0.0051	0.0053	0.0039	0.0051	0.0042	0.0033	0.0023	0.0015

Compute the covariance of age and Aboriginal, noting that  $E(Abo)=0.0306$  and  $E(AGEGRP)=12.0$ .  $Cov(Abo,AGEGRP)= E[(Abo-E(Abo))(Age-E(Age))]$ .

The expectation is computed using the probabilities given in the cross tab.

AGEGRP and the aboriginal dummy have negative covariance of -0.00566.

Suppose we ran a regression of wages on dummies for EEGROUP and a single variable called age which gives the age of the person. Would this regression suffer from a multicollinearity problem?

There is only a little correlation, so there is only a little collinearity problem.

18. Consider the following table of means of wages by sex and Employment Equity Group for Canadian-born people aged 25-64 with nonmissing wages > 100 in Vancouver

Descriptive Statistics for WAGES  
 Categorized by values of SEX and EEGROUP  
 Date: 11/16/10 Time: 10:19  
 Sample: 1 56529 IF AGEGRP > 8 AND AGEGRP < 17 AND WAGES > 100 AND  
 CITIZEN = 1 AND WAGES < 8000000  
 Included observations: 13216

Mean Std. Dev. Obs.		EEGROUP			
		1	2	3	All
1		38514.30	39734.27	31705.19	38399.38
		30035.05	31343.37	30444.25	30200.86
		5660	626	222	6508
SEX 2		64043.52	53484.13	40185.79	62339.55
		78405.27	65537.10	29223.81	76427.61
		5856	669	183	6708
All		51496.16	46837.48	35537.17	50550.61
		61089.95	52335.55	30159.74	59640.09
		11516	1295	405	13216

Are wages independent of EEGROUP?

For wages to be independent of EEGROUP, their conditional distribution would have to be the same for each value of EEGROUP.

The average wage is lower for EEGROUP=3 for each sex and for both sexes together, so wages are not independent of EEGROUP (even if you condition on sex).

What is the conditional mean of wages given EEGROUP?

*These are the bottom set of means in the Table*

51496.16 46837.48 35537.17

Are wages homoskedastic over EEGROUP and SEX? Are wages homoskedastic over EEGROUP given SEX?

*Skedasticity is about the variance of wages given EEGROUP and SEX.*

*For women (sex=1), the standard deviation (square root of variance) of wages is about 30,000, so pretty homoskedastic.*

*For men (sex=2), the standard deviation is much lower for Aboriginals, so pretty heteroskedastic.*

*So not homoskedastic over both EEGROUP and SEX, but might be homoskedastic given sex=1.*

Suppose we regress wages on a constant, 2 dummies for EEGROUP and a dummy for male: would this regression be misspecified in terms of its functional form? If so, how so? Is this regression misspecified in terms of the variables included?

*Is a the effect of EEGROUP on mean wages the same for men and women? For women it looks to be about \$8000 between Aboriginals and non-Aboriginals. For men, it looks to be at least \$11,000. Thus, the regression would be misspecified.*

Suppose we ran a regression of wages on a constant, 2 dummies for EEGROUP, a dummy for male and 2 interaction terms between EEGROUP and male. What would the coefficients be in this estimated regression?

*Here we have 6 coefficients in the model, so they would capture all the variation in the 6 means reported in the Table. Let the EEGROUP dummies be VISMIN and ABO. The constant would be white women (38514.30), the VISMIN dummy would be the difference between white women and vismin women (39734.27-38514.30), the ABO dummy would be the difference between white women and Aboriginal women (31705.19-38514.30). The male dummy would be the difference between white women and white men (64043.52-38514.30). The EEGROUP\*male interactions would pick up the differences between visible minority women and visible minority men and between Aboriginal women and Aboriginal men.*