Econometrics I
Midterm Exam. II
Dec. 13, 2002

1. (20%) State with brief reasons whether the following statements are true, false, or uncertain.

   (a) (5%) With the fact that $(A^{-1})' = (A')^{-1}$, if A is symmetric, then A^{-1} is also symmetric.

   (b) (5%) If we drop an explanatory variable from a regression model, both R^2 and $\bar{R}^2$ (adjusted R^2) will decrease.

   (c) (5%) For a simple regression model without constant term, $Y_i = \beta X_i + u_i$, the OLS estimator of β is $\frac{\bar{Y}}{\bar{X}}$, where $\bar{Y}$ and $\bar{X}$ are the average of Y_i and X_i respectively.

   (d) (5%) If A and B are two $m \times m$ matrices, therefore both AB and BA exist. Then AB=BA.

2. (15%) The following regression is estimated as a production function:

$$\ln Q = 1.37 + 0.632 \ln K + 0.452 \ln L\,, \quad R^2 = 0.38$$

   where $\text{Var}(\hat{\beta}_K) = 0.160$, $\text{Var}(\hat{\beta}_L) = 0.045$, $\text{Cov}(\beta_K, \beta_L) = 0.0225$. The sample size is 100. Test the following hypotheses at the 5% level of significance.

   (a) (7%) $\beta_K = \beta_L$.

   (b) (8%) There are constant returns to scale.

3. (25%) Let the multiple regression model be written as

$$\underset{n\times 1}{y} = \underset{n\times K}{X}\ \underset{K\times 1}{\beta} + \underset{n\times 1}{u}$$

   where y is the column vector of the dependent variable, X is the matrix of explanatory variables, β is the column vector of the K coefficients and u is the column vector of the error terms. Assume that $\text{E}(u) = 0$ and the covariance matrix of u is $\text{E}(uu') = \sigma^2 I$. It is known that the OLS estimator of β is $\hat{\beta} = (X'X)^{-1}X'Y$. Let $\hat{\beta}^* = \hat{\beta} + Cy$ be another linear estimator of β, where C is a $K \times n$ matrix.

   (a) (5%) For $\hat{\beta}^*$ to be unbiased, what restriction do we need to impose on C?

   (b) (10%) Derive the covariance matrix of $\hat{\beta}^*$, $\text{Var}(\hat{\beta}^*)$?

   (c) (10%) Compare $\text{Var}(\hat{\beta}^*)$ with $\text{Var}(\hat{\beta})$. What theorem can we have from the result of comparing the covariance matrix of the two estimators.

4. (10%) Suppose we have the following model,

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i * X_i) + u_i$$

   where Y_i is the annual salary of a worker, X_i is the years of working experience, and D_i is dummy variable with

$$\begin{aligned} D_i &= 1 \text{ if male} \\ &= 0 \text{ if female} \end{aligned}$$

(a) (5%) What is the difference of male and female average salaries in terms of the regression coefficient and other variables?

(b) (5%) Suppose the dummy variable is defined as $D_i = 1$ if female and $D_i = -1$ if male, then what is the difference of male and female average salaries with 10 years of working experience.

5. (30%) You are given a data set consists of the following 3 variables (1) $schyr$: one's years of schooling (2) sex: 1 for male and 0 for female (3) $fschyr$: father's years of schooling.

(a) (5%) Write down the necessary Stata commands to create a new dummy variable c which indicates that an individual has at least 16 years of schooling (college education).

(b) (5%) When you regress c on sex and the constant term, what is the Stata command for doing this?

(c) (5%) you have the following result from (b),

$$c = .2622419 - .0345699\, sex$$
$$(.0073647)(.0099553)$$

with standard errors in the parentheses. What is the proportion of male who has a college education? What is the proportion of female who has a college education?

(d) (5%) Is man more or less likely than woman to have a college education? Is it significant at 5% level of significance?

Now you add $fschyr$ and the interaction between sex and $fschyr$ as the explanatory variables and have the following results.

Source	SS	df	MS
Model	130.360821	3	43.4536071
Residual	1248.30531	7484	.166796541
Total	1378.66613	7487	.184141329

Number of obs	=	7488
F(3, 7484)	=	260.52
Prob> F	=	0.0000
R-squared	=	0.0946
Adj R-squared	=	0.0942
Root MSE	=	.40841

c	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf. Interval]	
sex	.0030215	.0256093	0.12	0.906	$-.0471799$	.053223
$fschyr$	.0387226	.001979	19.57	0.000	.0348432	.042602
$sex * fschyr$	$-.0038253$	.0026601	-1.44	0.150	$-.0090398$	.0013893
cons.	$-.0856444$	.0191131	-4.48	0.000	$-.1231115$	$-.0481774$

(e) (5%) Draw the regression lines of c on $fschyr$ for male and female separately.

(f) (5%) Is the effect of father's schooling on c significantly different for son and daughter?

Suggested Answers for Midterm Exam. II

Dec. 20, 2002

1a True. Since A is symmetric, then $A' = A$. $(A^{-1})' = (A')^{-1} = A^{-1}$, so A^{-1} is also symmetric.

1b False. R^2 will definitely decrease but $\bar{R}^2$ will not necessarily decrease since $\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-K}$, $\frac{\partial \bar{R}^2}{\partial K} = \frac{n-1}{n-K}\frac{\partial R^2}{\partial K} - (1 - R^2)\frac{n-1}{(n-K)^2} \gtrless 0$

1c True. To minimize $\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta} X_i)^2$, the first order condition is $\sum_{i=1}^n (-2)(Y_i - \hat{\beta} X_i) X_i = 0$. Therefore, the OLS estimator of β is $\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$.

1d False. In general $AB \neq BA$.

2a $\text{Var}(\beta_K - \beta_L) = \text{Var}(\hat{\beta}_K) + \text{Var}(\hat{\beta}_L) - 2\text{Cov}(\beta_K, \beta_L) = 0.16 + 0.045 - 2 * 0.0225 = 0.16$. $t = \frac{0.632 - 0.452}{\sqrt{0.16}} = \frac{0.18}{0.4} = 0.45$, insignificant.

2b $\text{Var}(\beta_K + \beta_L) = \text{Var}(\hat{\beta}_K) + \text{Var}(\hat{\beta}_L) + 2\text{Cov}(\beta_K, \beta_L) = 0.16 + 0.045 + 2 * 0.0225 = 0.25$. $t = \frac{0.632 + 0.452 - 1}{\sqrt{0.25}} = \frac{0.084}{0.5} = 0.168$, insignificant.

3a $\text{E}[\hat{\beta}^*] = \text{E}[\hat{\beta} + CX\beta + u] = \beta + CX\beta$. For $\hat{\beta}^*$ to be unbiased, we need $CX = 0$.

3b

$$
\begin{aligned}
\hat{\beta}^* &= \beta + (X'X)^{-1}X'u + Cu \\
\text{Var}(\hat{\beta}^*) &= \text{E}[(\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)'] \\
&= \text{E}\{[(X'X)^{-1}X'u + Cu][(X'X)^{-1}X'u + Cu]'\} \\
&= \text{E}\{[(X'X)^{-1}X'u + Cu][u'XX'X)^{-1} + u'C']\} \\
&= \text{E}[(X'X)^{-1}X'uu'XX'X)^{-1} + Cuu'XX'X)^{-1} + (X'X)^{-1}X'uu'C' + Cuu'C'] \\
&= \sigma^2(X'X)^{-1} + 0 + 0 + \sigma^2 CC' \geq \text{Var}(\hat{\beta})
\end{aligned}
$$

3c $\text{Var}(\hat{\beta}^*) \geq \text{Var}(\hat{\beta})$. This is the Gauss-Markov theorem, the OLS estimator of β–$\hat{\beta}$ is the most efficient estimator among the class of the linear unbiased estimators.

4a

$$
\begin{aligned}
\text{E}(Y_i|D_i = 1) &= \beta_1 + \beta_2 + \beta_3 X_i + \beta_4 X_i \\
\text{E}(Y_i|D_i = 0) &= \beta_1 + \beta_3 X_i \\
\text{E}(Y_i|D_1 = 1) - \text{E}(Y_i|D_i = 0) &= \beta_2 + \beta_4 X_i
\end{aligned}
$$

4b

$$
\begin{aligned}
\text{E}(Y_i|D_i = -1, X_i = 10) &= \beta_1 - \beta_2 + \beta_3 * 10 - \beta_4 * 10 \\
\text{E}(Y_i|D_i = 1, X_i = 10) &= \beta_1 + \beta_2 + \beta_3 * 10 + \beta_4 * 10 \\
\text{E}(Y_i|D_1 = -1, X_i = 10) - \text{E}(Y_i|D_i = 1, X_i = 10) &= -2\beta_2 - 20\beta_4
\end{aligned}
$$

5a generate c=1 if schyr$>=$ 16
replace c=0 if schyr==.

5b regress c sex

5c male: 26.2-3.5=22.7%, female: 26.2%.

5d $t = \frac{-0.0345699}{0.0099553} = -3.47$. Man is significantly less likely to have a college education.

5e

	son	daughter
intercept	-0.083	-0.086
slope	0.035	0.039

5f

Since the coefficient of $sex * fschyr$ is not significant, the effect of father schooling on c is not significant different for son and daughter.