Final exam for BUEC 333 D200, Spring 2012
December 13, 2012, 12:00-15:00, SSCB 9201

- This final consists of 13 questions.
- For **every sub**question (e.g. 4(c)) , use at most 3 lines (not sentences) of text to answer. This does not include math, computations, and graphs/drawings. **If you write more than 3 lines of answer, you will receive 0 points for that subquestion.**
- Statistical tables can be found after question 13.
- Allowed on your desk:
  - non-graphical calculator;
  - pens, pencils, but **no pencil cases or erasers**;
  - this exam;
  - an answer sheet;
  - That's all!

Questions

1. Consider the joint probability distribution of the random variables X_1 and X_2 given in Table 1.

   (a) What is the probability that $X_1 = 0$ given that $X_2 = 2$?
   In other words, what is $P(X_1 = 0 | X_2 = 2)$?

   (b) What is the conditional expectation of X_1 given that $X_2 = 3$? In other words, what is $E(X_1 | X_2 = 2)$?

2. Let X be a random variable, with expectation $E[X] = \mu$, variance $\text{Var}[X] = \sigma^2$. Compute $\text{Var}[\mu]$, the variance of μ.

3. This question is about the stochastic error term typically denoted by ϵ.

   (a) Why do we typically include a stochastic error term in our regression equation?

   (b) What is the difference between a stochastic error term and a residual?

4. This is a back-of-chapter-2 exercise. Consider the following estimated regression equation that describes the relationship between a student's weight and height:

$$\widehat{WEIGHT} = 103.40 + 6.38\, HEIGHT$$

   Now, an additional variable is included, is MAIL, a student's post office box number. Obviously, this is a nonsensical variable that is unrelated to a student's weight. The new estimated regression equation is

$$\widehat{WEIGHT} = 102.35 + 6.36\, HEIGHT + 0.02\, MAIL$$

   (a) Someone's weight has nothing to do with their post office box number. Still, the R^2 went up from 0.74 to 0.75. How is this possible?

   (b) On the other hand, the adjusted R-squared $\bar{R}^2$ went down from 0.73 to 0.72. Explain how it is possible that the R^2 can go up while the $\bar{R}^2$ goes down.

   (c) If the post office box number is not related to a student's weight, should the estimated coefficient not be equal to 0? How could it be that it is 0.02?

5. Let X be a random variable with mean $\text{E}[X] = \mu$. We have a random sample of observations on X, namely $(X_1, \cdots, X_n)$. Let $\bar{X}$ denote the sample average, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Now, let $\widetilde{X} = \frac{1}{n-2}\sum_{i=1}^{n} X_i$ be the scaled sample average. Is $\widetilde{X}$ unbiased?

6. Explain what the p-value is. Based on the p-value, when do you reject the null hypothesis? Assume the case of a one-sided t-test? Draw a picture to illustrate your answer.

	$X_2 = 1$	$X_2 = 2$	$X_2 = 3$
$X_1 = 0$	0.2	0.1	0.1
$X_1 = 1$	0.3	0.3	0

Table 1: Joint probability distribution of X_1 and X_2.

7. This question is about doing OLS without a constant term. Using a picture, show what happens when you suppress the constant term when $\beta_0 > 0$ and $\beta_1 > 0$, and your model is $Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$. Do you expect $\hat{\beta}_1$ to have a positive or negative bias? Explain, and use a graph to illustrate your answer.

8. Let D_i be a dummy variable. Consider the model

$$\log(Y_i) = \beta_0 + \beta_1 X_i + \beta_2 D_i + \epsilon_i.$$

How do you interpret β_2? Use the fact that the dependent variable is in logs.

9. Let D_i be a dummy variable. Consider the model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + \epsilon_i.$$

Draw a graph and use it to provide the interpretation of the coefficient β_3.

10. Let W_i be an individual's wage, let $EDUC_i$ be their education, and let F_i be a dummy variable that equals 1 if and only if that person is female. Furthermore, let M_i be a dummy variable that equals 1 if and only if that person is male. Consider the following model:

$$W_i = \beta_0 + \beta_1 EDUC_i + \beta_2 F_i + \beta_3 M_i + \epsilon_i.$$

   (a) What is the problem with using OLS to estimate the parameters in this regression equation?

   (b) How do you solve this problem?

11. Your regression output tells you that the R^2 is high, but you notice that all of the t-ratios are low.

   (a) What does this signal?

   (b) Explain why it signals this.

   (c) What can you do to remedy this?

12. Consider two models of Soviet defense spending before the breakup of the Soviet Union. Soviet defense spending is believed to be a function of U.S. Defense spending and Soviet GNP. It may be a function of the ratio of US to Soviet warheads, too. The dependent variable is log(SDH)t) (Soviet defense expenditure in year t in billions of Rubles), and possible explanatory variables are:

   - USD$_t$: U.S. defense expenditures in year t (billions of U.S. dollars)
   - SY$_t$: Soviet GNP in year t (billions of Rubles)
   - SP$_t$: ratio of the number of USSR to U.S nuclear warheads in year t

   The two models are:

$$\log(SDH_t) = \beta_0 + \beta_1 \log(USD_t) + \beta_2 \log(SY_t) + \beta_3 \log(SP_t) + \epsilon_t, \quad (1)$$
$$\log(SDH_t) = \beta_0 + \beta_1 \log(USD_t) + \beta_2 \log(SY_t) + \epsilon_t. \quad (2)$$

   You estimate both models using R, which gives you the following estimates and standard errors (SE):

	Model (1)		Model (2)	
	Estimate	SE	Estimate	SE
$\hat{\beta}_1$	0.056	0.074	0.105	0.073
$\hat{\beta}_2$	0.969	0.065	1.066	0.038
$\hat{\beta}_3$	0.057	0.032		
N	25		25	
$\bar{R}^2$	0.979		0.977	
DW	0.49		0.43	

Answer the following question:

(a) Based on common sense/theory, would you include the additional variable?

(b) What about the other three criteria that we use to assess whether a variable is irrelevant/omitted? Explain.

(c) On the basis of (a) and (b), which model do you prefer?

(d) Test both equations for positive first-order serial correlation. Use a test at the 5-percent level. Make sure to properly formulate the null and alternative hypothesis before you perform the test. What do you conclude?

13. Say that you have a data set called "unionData" that contains two variables "lwage" and "union". To get the sample average of "lwage", you give the command "mean(unionData$lwage)".

(a) What is the command to run a regression of "lwage" on "union"?

(b) What do you expect for the sign of "union"?

TABLE B-1 CRITICAL VALUES OF THE t-DISTRIBUTION

Degrees of Freedom	Level of Significance One Sided: 10% Two Sided: 20%	5% 10%	2.5% 5%	1% 2%	0.5% 1%
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
(Normal) ∞	1.282	1.645	1.960	2.326	2.576

Source: Reprinted from Table IV in Sir Ronald A. Fisher, *Statistical Methods for Research Workers*, 14th ed. (copyright © 1970, University of Adelaide) with permission of Hafner, a Division of the Macmillan Publishing Company, Inc.

TABLE B-4 CRITICAL VALUES OF THE DURBIN–WATSON TEST STATISTICS D_L AND D_U: 5 PERCENT ONE-SIDED LEVEL OF SIGNIFICANCE (10 PERCENT TWO-SIDED LEVEL OF SIGNIFICANCE)

n	k′ = 1		k′ = 2		k′ = 3		k′ = 4		k′ = 5		k′ = 6		k′ = 7	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.81	1.75	0.69	1.97	0.56	2.21	0.45	2.47	0.34	2.73
16	1.11	1.37	0.98	1.54	0.86	1.73	0.73	1.93	0.62	2.15	0.50	2.39	0.40	2.62
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.66	2.10	0.55	2.32	0.45	2.54
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06	0.60	2.26	0.50	2.46
19	1.18	1.40	1.07	1.53	0.97	1.68	0.86	1.85	0.75	2.02	0.65	2.21	0.55	2.40
20	1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99	0.69	2.16	0.60	2.34
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96	0.73	2.12	0.64	2.29
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94	0.77	2.09	0.68	2.25
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92	0.80	2.06	0.72	2.21
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90	0.84	2.04	0.75	2.17
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89	0.87	2.01	0.78	2.14
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88	0.90	1.99	0.82	2.12
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.00	1.86	0.93	1.97	0.85	2.09
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85	0.95	1.96	0.87	2.07
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84	0.98	1.94	0.90	2.05
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	1.00	1.93	0.93	2.03
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83	1.02	1.92	0.95	2.02
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82	1.04	1.91	0.97	2.00
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81	1.06	1.90	0.99	1.99
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.14	1.81	1.08	1.89	1.02	1.98
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80	1.10	1.88	1.03	1.97
36	1.41	1.52	1.35	1.59	1.30	1.65	1.24	1.73	1.18	1.80	1.11	1.88	1.05	1.96
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80	1.13	1.87	1.07	1.95
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.20	1.79	1.15	1.86	1.09	1.94
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79	1.16	1.86	1.10	1.93
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	1.18	1.85	1.12	1.93
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78	1.24	1.84	1.19	1.90
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77	1.29	1.82	1.25	1.88
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.37	1.77	1.33	1.81	1.29	1.86
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.37	1.81	1.34	1.85
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77	1.40	1.81	1.37	1.84
70	1.58	1.64	1.55	1.67	1.53	1.70	1.49	1.74	1.46	1.77	1.43	1.80	1.40	1.84
75	1.60	1.65	1.57	1.68	1.54	1.71	1.52	1.74	1.49	1.77	1.46	1.80	1.43	1.83
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.48	1.80	1.45	1.83
85	1.62	1.67	1.60	1.70	1.58	1.72	1.55	1.75	1.53	1.77	1.50	1.80	1.47	1.83
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.52	1.80	1.49	1.83
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78	1.54	1.80	1.51	1.83
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.55	1.80	1.53	1.83

Source: N. E. Savin and Kenneth J. White. "The Durbin–Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors," *Econometrica,* November 1977, p. 1994. Reprinted with permission.

Note: n = number of observations, k′ = number of explanatory variables excluding the constant term. We assume the equation contains a constant term and no lagged dependent variables (if so see Table B-7).

TABLE B-5 CRITICAL VALUES OF THE DURBIN–WATSON TEST STATISTICS OF D_L AND D_U: 2.5 PERCENT ONE-SIDED LEVEL OF SIGNIFICANCE (5 PERCENT TWO-SIDED LEVEL OF SIGNIFICANCE)

n	k′ = 1		k′ = 2		k′ = 3		k′ = 4		k′ = 5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
16	0.98	1.24	0.86	1.40	0.75	1.59	0.64	1.80	0.53	2.03
17	1.01	1.25	0.90	1.40	0.79	1.58	0.68	1.77	0.57	1.98
18	1.03	1.26	0.93	1.40	0.82	1.56	0.72	1.74	0.62	1.93
19	1.06	1.28	0.96	1.41	0.86	1.55	0.76	1.72	0.66	1.90
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
21	1.10	1.30	1.01	1.41	0.92	1.54	0.83	1.69	0.73	1.84
22	1.12	1.31	1.04	1.42	0.95	1.54	0.86	1.68	0.77	1.82
23	1.14	1.32	1.06	1.42	0.97	1.54	0.89	1.67	0.80	1.80
24	1.16	1.33	1.08	1.43	1.00	1.54	0.91	1.66	0.83	1.79
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
26	1.19	1.35	1.12	1.44	1.04	1.54	0.96	1.65	0.88	1.76
27	1.21	1.36	1.13	1.44	1.06	1.54	0.99	1.64	0.91	1.75
28	1.22	1.37	1.15	1.45	1.08	1.54	1.01	1.64	0.93	1.74
29	1.24	1.38	1.17	1.45	1.10	1.54	1.03	1.63	0.96	1.73
30	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
31	1.26	1.39	1.20	1.47	1.13	1.55	1.07	1.63	1.00	1.72
32	1.27	1.40	1.21	1.47	1.15	1.55	1.08	1.63	1.02	1.71
33	1.28	1.41	1.22	1.48	1.16	1.55	1.10	1.63	1.04	1.71
34	1.29	1.41	1.24	1.48	1.17	1.55	1.12	1.63	1.06	1.70
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
36	1.31	1.43	1.26	1.49	1.20	1.56	1.15	1.63	1.09	1.70
37	1.32	1.43	1.27	1.49	1.21	1.56	1.16	1.62	1.10	1.70
38	1.33	1.44	1.28	1.50	1.23	1.56	1.17	1.62	1.12	1.70
39	1.34	1.44	1.29	1.50	1.24	1.56	1.19	1.63	1.13	1.69
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.63	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
55	1.45	1.52	1.41	1.56	1.37	1.60	1.33	1.64	1.30	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
65	1.49	1.55	1.46	1.59	1.43	1.62	1.40	1.66	1.36	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
75	1.53	1.58	1.50	1.61	1.47	1.64	1.45	1.67	1.42	1.70
80	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
85	1.56	1.60	1.53	1.63	1.51	1.65	1.49	1.68	1.46	1.71
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
95	1.58	1.62	1.56	1.65	1.54	1.67	1.52	1.69	1.50	1.71
100	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72

Source: J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression," *Biometrika,* Vol. 38, 1951, pp. 159–171. Reprinted with permission of the *Biometrika* trustees.
Note: n = number of observations, k′ = number of explanatory variables excluding the constant term. *It is assumed that the equation contains a constant term and no lagged dependent variables (if [illegible] see Table B-7).*