STOCKHOLMS UNIVERSITET
Statistiska institutionen
Oskar Gustafsson

WRITTEN RE-EXAM, ECONOMETRICS I
2022-01-03

Time for examination: 8.00-13.00

Allowed tools: Pocket calculator, own formula sheet (1 double-sided A4 page), Course text-book: Wooldridge, J.M. *Introductory Econometrics - a Modern Approach (any edition)*
Note that no formula sheet will be provided.

The exam consists of 4 independent problems. Well motivated and clear solutions are required for full scoring on a problem. Don't forget to state any necessary assumptions or conditions where needed.
Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description. Answers may be given in English or Swedish.

Good luck!

Problem 1. (25 points)
Indicate which alternative that is correct. Answering more than one alternative result in 0 points on the sub-question. No motivation is required.

1. If $\hat{\beta}_1$ is consistent for β_1 we have that:
   (a) $\hat{\beta}_1 = \beta_1$ in a sample.
   (b) $E[\hat{\beta}_1] = \beta_1$.
   (c) $\hat{\beta}_1$ converges in probability to β_1 as $n \to \infty$.
   (d) $\text{var}(\hat{\beta}_1)$ converges to a constant greater than 0 as $n \to \infty$.

2. Which the following is **not** related to multicolinarity problems:
   (a) We have included a regressor that is uncorrelated with our dependent variable.
   (b) Some of our independent variables are highly correlated.
   (c) We have a high variance inflation factor.
   (d) We get a high R^2 and high variance in our parameter estimates for the slopes.

3. Assuming that MLR 1-6 are fulfilled for the regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, we have (all else equal) that $\text{var}(\hat{\beta}_1)$:
   (a) decreases if $\text{var}(y)$ increases.
   (b) decreases if $\text{var}(x_1)$ increases.
   (c) increases with the R^2 of the model.
   (d) approaches a positive constant as $n \to \infty$.

4. Consider a Poisson regression with $\log \lambda = \log E[y|x] = 1 + 0.2x$. What is (approximately) the probability that $y = 3$ if $x = 1.93$?
   (a) 0.1
   (b) 0.2
   (c) 0.3
   (d) 0.4

5. 2SLS is more efficient than IV-regression when:
   (a) we have several exogenous independent variables.
   (b) we have correlation between our instrument and the error term.
   (c) y is measured with error.
   (d) we have more than one instrumental variable for the endogenous variable.

Problem 2. (25 points)

Consider the simple linear regression model:

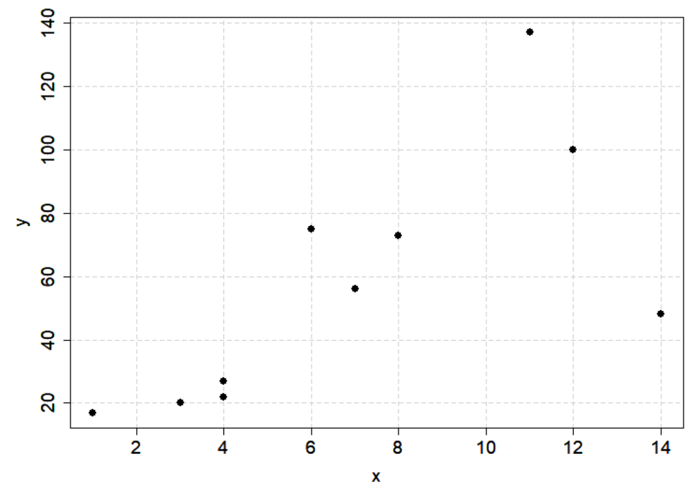
$$y = \beta_0 + \beta_1 x + u,$$

where we have observed the data in Table 1

Table 1: Data

obs.	x	y	$\hat{y}$	$\hat{u}$	$\hat{u}^2$
1	1	17			2.25
2	3	20			132.25
3	4	22			256
4	4	27			121
5	6	75			576
6	7	56			
7	8	73			
8	11	137			
9	12	100			
10	14	48			
sum	70	575			

Note: Simulated data.



Questions:

a) Use the data in Table 1 to calculate $\hat{\beta}_1$ [*Hint: you may use the fact that* $\hat{\beta}_0 = 12$]

b) Calculate SST, SSR, SSE, R^2 and $\hat{\sigma}^2$.

c) Construct a 95% confidence interval for β_1 (assuming that MLR 1-6 hold).

d) We may suspect that the data is heteroscedastic. Test this using the Breuch-Pagan test. *[Hint: since we use a simple linear regression model it is enough to test whether our single x is correlated with the squared residuals. You may use that the intercept and residual standard errors of the Breush-Pagan regression are given by $\hat{\beta}_0 = -621.78$ and $\hat{\sigma}^2 = 918.6$ respectively to test this].*

e) No matter the result of your test in (c), describe how we can use WLS to estimate the model parameters. Also scale the data in Table 1 appropriately using the relation $\sigma_i^2 = \sigma^2 h_i(x_i) = \sigma^2 x_i^2$.

Problem 3. (25 points)

(a) Determine which of the following models that are linear in the parameters, has constant partial effect or both

$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 \log x_2 + u \tag{1}$$
$$y = \beta_0 + \beta_1 x_1 + \beta_1 \beta_2 x_2 + u \tag{2}$$
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u \tag{3}$$
$$y = \beta_0 + \beta_1 x_1 + (\beta_2 + \beta_1) x_2^2 + u \tag{4}$$
$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + u \tag{5}$$

(b) Interpret β_1 in model (3) and β_2 in model (1) from (a).

(c) Consider model (1) and assume that all our assumptions (MLR.1-MLR.6) are fulfilled. Name at least three important properties that the OLS estimators for β_0, β_1 and β_2 has.

(d) Let's say that you are interested in testing the hypothesis that $\delta = \beta_1 + \beta_2 = 3$ in model (1). Give a detailed description of how you can construct a 95% confidence interval for δ (still assuming that MLR 1-6 are fulfilled).

Problem 4. (25 points)

A stock market analyst measures companies "quarterly profit surprises" by:

$$x_1 = \frac{\text{profit per stock} - \text{expected profit per stock}}{\text{price per stock}}.$$

The analyst is interested in whether a positive "profit surprise" increase the probability that the stock price will rise in the coming week. She also decided to control for *the average of the absolute values of previously reported profit surprises*, x_2, as an additional regressor. After collecting the data and running a logistic regression the analyst obtain the following:

$$P(y = 1|\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)} = \frac{\exp(1 + 70x_1 - 20x_2)}{1 + \exp(1 + 70x_1 - 20x_2)} \tag{6}$$

where $y = \begin{cases} 1, & \text{if the stock price went up} \\ 0, & \text{if the stock price went down} \end{cases}$

Table 2: Stock Data

obs.	x_1	x_2	y
1	−0.012	0.045	1
2	−0.001	0.040	1
3	0.004	0.010	1
4	0.014	0.030	1
5	0.023	0.050	1
6	−0.017	0.050	0
7	−0.007	0.035	0
8	0.008	0.020	0
Mean	0.002	0.020	0.6

Note: Consider these data as a subset of the data that was used to estimate model in (6).

Queastions:

(a) According to her model, what is the probability that the stock of a company with $x_2 = 0.05$ will increase if they report a profit surprise of 3%?

(b) What is the probability that the stock price of the same companies increase if they report a profit surprise of −1%?

(c) Do your answers in (a) and (b) make economic sense? Explain.

(d) Derive the expression for the partial effect on $P(y = 1|x_1, x_2)$ for a change in x_1.

(e) Calculate the *average partial effect* (APE) and the *partial effect at the average* PEA, using your answer in (d) and Table 2. Since Table 2 does not contain all the data, you should actually calculate the local partial effects associated with the sub-sample.

(f) What would be the implications if the analyst instead decides to use the *linear probability model* (LPM) or the *probit* model? Which should she prefer, and why?

TABLE G.2 Critical Values of the *t* Distribution

		Significance Level				
1-Tailed:		.10	.05	.025	.01	.005
2-Tailed:		.20	.10	.05	.02	.01
Degrees of Freedom	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
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical value for a two-tailed test with large (>120) *df* is 1.96.

Source: This table was generated using the Stata® function invttail.