SUMMER TERM 2020

ECON0019: QUANTITATIVE ECONOMICS AND ECONOMETRICS

TIME ALLOWANCE: 3 hours

Answer ALL TWO questions from Part A and answer ONE question from Part B.

Questions in Part A carry 60 per cent of the total mark and questions in Part B carry 40 per cent of the total. Tables for the normal and F-distribution are at the end of the examination paper.

In cases where a student answers more questions than requested by the examination rubric, the policy of the Economics Department is that the student's first set of answers up to the required number will be the ones that count (not the best answers). All remaining answers will be ignored.

PART A

Answer all questions from this section.

A.1 You have randomly sampled n individuals whom you follow over T time periods. For individual i $(= 1, ..., n)$ you observe (y_{it}, x_{it}), $t = 1, ..., T$, which satisfies

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}, \quad t = 2, ..., T. \tag{1}$$

(a) Show that

$$\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta u_{it}, \quad i = 1, ..., n, \; t = 2, ..., T. \tag{2}$$

Discuss the advantages and disadvantages of using eq. (2) instead of eq. (1) for estimation and inference.

(b) Write up the sum of squared residuals (SSR) for (2) given your sample. Suppose here and in the following that $\sum_{i=1}^{n} \sum_{t=1}^{T} (\Delta x_{it})^2 > 0$ in your sample. Show that the minimizer of SSR is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \Delta x_{it} \Delta y_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{T} (\Delta x_{it})^2}.$$

(c) Suppose that

$$u_{it} = u_{it-1} + e_{it}$$

where $E[e_{it} | x_{i1}, ..., x_{iT}] = 0$, $t = 2, ..., T$. Show that for any i, t,

$$E[\Delta u_{it} | X_1, ..., X_n] = 0,$$

where $X_i = (\Delta x_{i2}, ..., \Delta x_{iT})$, $i = 1, ..., n$. Use this in turn to show that

$$E[\hat{\beta}_1 | X_1, ..., X_n] = \beta_1.$$

(d) Suppose furthermore that

$$\begin{aligned} E\left[e_{it}^2 | x_{i1}, ..., x_{iT}\right] &= \sigma^2, \\ E\left[e_{is}e_{it} | x_{i1}, ..., x_{iT}\right] &= 0, \quad s \neq t. \end{aligned}$$

Demonstrate that

$$\text{Cov}\left(\Delta u_{is}, \Delta u_{it} | X_1, ..., X_n\right) = \begin{cases} \sigma^2, & s = t \\ 0, & s \neq t \end{cases}.$$

Use this in turn to derive the following variance formula,

$$\text{Var}(\hat{\beta}_1 | X_1, ..., X_n) = \frac{\sigma^2}{\sum_{i=1}^n \sum_{t=1}^T (\Delta x_{it})^2}.$$

(e) Assume that $E\left[\sum_{t=1}^T (\Delta x_{it})^2\right] > 0$. Show consistency of $\hat{\beta}_1$.

A.2 An extension of the Solow growth model, that includes human capital in addition to physical capital, suggests that investment in human capital (education) will increase the wealth of a nation (per capita income). To test this hypothesis, you collect data for 104 countries and perform the following regression:

$$\underset{}{\widehat{relinc}} = \underset{(0.079)}{0.046} - \underset{(2.238)}{5.869}gpop + \underset{(0.294)}{0.738}sk + \underset{(0.010)}{0.055}educ, \quad (3)$$

with $R^2 = 0.775$, standard error of residual $SER = 0.1377$, and heteroskedasticity-robust standard errors reported in parentheses. Here, *relinc* is GDP per worker relative to the United States, *gpop* is the average population growth rate, 1980 to 1990, *sk* is the average investment share of GDP from 1960 to 1990, and *educ* is the average educational attainment in years for 1985.

(a) Discuss the implications and validity of each of the following assumptions in the context of the above regression:

i. Data is i.i.d.

ii. $E[u|gpop, sk, educ] = 0$ where u is the regression error.

In the following we will assume that (i)-(ii) are satisfied together with other relevant technical assumptions.

(b) Interpret the results and indicate whether or not the coefficients are significantly different from zero. Do the coefficients have the expected sign?

(c) To test for equality of the coefficients between the OECD and other countries, you introduce a binary variable (*oecd*), which takes on the value of one for the OECD countries and is zero otherwise. You obtain the following regression estimates:

$$\begin{aligned} \widehat{relinc} \;=\; & \underset{(0.072)}{-0.068} - \underset{(2.271)}{0.063} gpop + \underset{(0.365)}{0.719} sk + \underset{(0.012)}{0.044} educ \\ & + \underset{(0.184)}{0.381} oecd - \underset{(5.366)}{8.038}(oecd \times gpop) - \underset{(0.768)}{0.430}(oecd \times sk) \\ & + \underset{(0.018)}{0.003}(oecd \times educ) \end{aligned} \qquad (4)$$

where $R^2 = 0.845$ and $SER = 0.116$. Write down the two regression functions, one for the OECD countries, the other for the non-OECD countries. Interpret any differences.

(d) In order to test (3) against (4), you compute the corresponding F-statistic which takes the value 6.76 in your sample. Write up the null hypothesis and its alternative that you are testing in terms of the population regression coefficients. What do you conclude?

(e) You decide to investigate further and estimate a restricted version of (4) where you enforce the same slopes across OECD and non-OECD countries, but allow their intercepts to differ. In this new regression, the t-statistic for *oecd* is 3.17. What is the p-value of the t-statistic? What do you conclude?

(f) Next, you test the model described in (e) against (4). The value of the corresponding F-statistic is 1.05. Do you accept or reject the null?

Looking at the tests in this and two previous questions, what is your overall conclusion?

PART B

Answer ONE question from this section.

B.1 Intergenerational mobility is related to several aspects. For example, theoretical studies have examined the repercussions of the transmission of preferences and attitudes from parents to children. Thomas Dohmen, Armin Falk, David Huffman and Uwe Sunde ("The Intergenerational Transmission of Risk and Trust Attitudes") use the German Socio-Economic Panel Study (SOEP) to empirically examine, among other things, the transmission of attitudes from parents to children and potential mechanisms for such transmission. Aside from comprehensive demographic information on all individuals in a given household, the survey contains a set of individual questions regarding risk attitudes (in 2004). (The authors also look at trust.) People were asked questions eliciting their willingness to take risks on an eleven-point scale. For these variables, zero (0) would correspond to 'completely unwilling to take risks' and the value ten (10) means that the person is 'completely willing to take risks.'

(a) One possible way to investigate the transmission of risk attitudes is to examine how parental characteristics (including their risk attitudes) relate to the probability that a child has a high score in terms of the risk attitude measure elicited on an 11-point scale as indicated above. To do this, generate a variable $D_i = 1$ if the child in household i has risk attitude measure equal to 6 or above and $D_i = 0$, otherwise. (While separate measures are available for both parents, to keep matters simple we focus here on a single measure for parents.) Taking R_i^P to be the parental score for that same measure in the household, suppose you are interested in the model:

$$D_i = \mathbf{1}(\beta_0 + \beta_1 R_i^P + U_i \geq 0).$$

Assuming that U_i follows a standard logistic distribution, write down the log-likelihood function for this estimation problem when you have N observations. How would you estimate the difference in the probability that $D_i = 1$ between a household where $R_i^P = 10$ and another one where $R_i^P = 0$?
Hint: The CDF for a standard logistic distribution is given by $F(x) = \exp(x)/(1 + \exp(x))$ *and its PDF is given by* $f(x) = \exp(x)/(1 + \exp(x))^2$.

(b) Because risk attitudes for children (R_i^C) and parents (R_i^P) are measured contemporaneously, the authors worry about 'reverse causality' where children's attitudes may be at least partly shaping parents' attitudes. To address this issue they estimate

$$R_i^C = \alpha_0 + \alpha_1 R_i^P + V_i,$$

using parental religion (Z_i) as an instrumental variable for R_i^P. Describe how you would implement the TSLS estimator in this context. Discuss the validity of the instrumental variable suggested in this context.

(c) The F-statistic for the first stage regression using the mother's risk attitudes as covariate in the main equation of interest and her religion as instrumental variable is 9.99. (The F-statistic when using father's risk attitudes and religion is 7.32.) Discuss the relevance of the instrumental variable.

(d) In a regression where the risk attitude for both mother and father are included individually as covariates in a multiple linear regression model, both coefficients on those variables are around 0.15 with standard errors at around 0.02 for each one of them. The TSLS estimates on the other hand, produce estimates for the coefficient on the mother's risk attitude at about 0.23 and for the coefficient on the father's risk attitude at about 0.02. (Religion for each parent is available as an intrumental variable for each of their risk attitude variables.) The standard error for those estimates are, in both cases, around 0.10. Why would you expect the standard errors for the IV estimates to be larger than the standard errors for the OLS estimates? Explain your answer.

(e) Imagine you had data on the risk attitude for successive generations of a single household and you want to estimate the regression

$$R_{G+1} = \alpha_0 + \alpha_1 R_G + V_{G+1},$$

where R_{G+1} and R_G are, once again, the risk attitudes in generation $G+1$ (child) and in generation G (parent). Assuming these are not measured contemporaneously so that the issues raised in item (b) are not present, are there conditions under which an OLS estimator is unbiased? Elaborate on your answer.

B.2 In "Excess Capacity and Policy Interventions: Evidence from the Cement Industry," Tetsuji Okazaki, Ken Onishi and Naoki Wakamori estimate the demand for cement in Japan using data on different regions across years. Their specification for the demand function is

$$\ln(Q_{mt}) = \alpha_P \ln(P_{mt}) + \alpha_X^\top X_{mt} + U_{mt},$$

where Q_{mt} is the quantity of cement demanded in region m and year t (from 1970 to 1995), P_{mt} is the price in that region and year and X_{mt} are year- and region-specific demand shifters. The Ordinary Least Squares (OLS) estimate for α, denoted by $\widehat{\alpha}_{P,\mathrm{OLS}}$, equals -0.07 with a standard error equal to 0.16.

(a) Explain why the above estimate for the slope coefficient (-0.07) cannot be directly interpreted as the price-elasticity of demand for cement.

(b) To produce cement, crushed limestone, cray and other minerals are mixed and put into a kiln to be heated. This process yields clinker, which is an intermediate cement product. In a final stage, the grinded clinker is mixed with gypsum, another intermediate input, to produce cement. The researchers then use the (log) price of gypsum as an instrumental variable for the (log) price of cement to estimate the price-elasticity of demand. The OLS regression of (log) cement prices on (log) gypsum prices (and X) yields a coefficient of 0.06 and the F-test statistic for the first stage equals 17.0. Discuss the exogeneity and relevance of this instrumental variable.

(c) To estimate the regression using the IV described above, the researchers use Two-Stage Least Squares and obtain an estimate for α, denoted $\widehat{\alpha}_{P,\text{TSLS}}$, equal to -1.11 with a standard error equal to 0.58. Describe in detail the TSLS procedure. Is it possible to test whether the IV is exogenous? Explain in detail. What if there were two instrumental variables? Explain.

(d) Suppose the researchers were also interested in examining the time series behaviour for the quantity of cement sold in a particular region in Japan on a given year, $\ln(Q_t)$. To do so, they obtain estimates for the following autoregressive model using data over various years for this region of Japan:

$$\ln(Q_t) = \alpha_0 + \alpha_1 \ln(Q_{t-1}) + \eta_t.$$

Would the OLS estimator be unbiased in this case? Under what assumptions would it be consistent?

(e) Suppose the researchers only observe whether Q_{mt} is larger or smaller than a given fixed threshold $\overline{Q}$ in a given year but otherwise observe prices and X. Let D_{mt} record whether $Q_{mt} > \overline{Q}$ ($D_{mt} = 1$) or not ($D_{mt} = 0$). While the regression

$$\ln(Q_{mt}) = \alpha_P \ln(P_{mt}) + \alpha_X^\top X_{mt} + U_{mt}$$

is no longer estimable, they are still able to estimate the model given by

$$D_{mt} = \begin{cases} 1 & \text{if } \beta_P \ln(P_{mt}) + \beta_X^\top X_{mt} + V_{mt} > \ln(\overline{Q}) \\ 0 & \text{if } \beta_P \ln(P_{mt}) + \beta_X^\top X_{mt} + V_{mt} \leq \ln(\overline{Q}) \end{cases}$$

Assume that the error term follows a standard normal distribution (i.e., $V_{mt} \sim \mathcal{N}(0,1)$) and write down the log-likelihood function for this model assuming that the data comprises of a random sample. If $U_{mt} \sim \mathcal{N}(0, \sigma^2)$ how are β_P and α_P related?

5 % Critical values for the F_{ν_1,ν_2} distribution

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	10	12	15	20	30	50	∞
1	161	199.	216.	225.	230.	234.	237.	239.	242.	244.	246.	248.	250.	252.	254.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.79	8.74	8.70	8.66	8.62	8.58	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.96	5.91	5.86	5.80	5.75	5.70	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.74	4.68	4.62	4.56	4.50	4.44	4.36
10	4.96	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.38	2.31	2.23	2.16	2.07	2.00	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.35	2.28	2.20	2.12	2.04	1.97	1.84
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.16	2.09	2.01	1.93	1.84	1.76	1.62
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.99	1.92	1.84	1.75	1.65	1.56	1.39
80	3.97	3.11	2.72	2.49	2.33	2.21	2.13	2.06	1.95	1.88	1.79	1.70	1.60	1.51	1.32
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.93	1.85	1.77	1.68	1.57	1.48	1.28
120	3.91	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.91	1.83	1.75	1.66	1.55	1.46	1.25
∞	3.85	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.83	1.75	1.67	1.57	1.46	1.35	1.00

NORMAL CUMULATIVE DISTRIBUTION FUNCTION ($Prob(z < z_a)$ where $z \sim N(0,1)$)

z_a	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995