Queen Mary University of London

January Examination Period 2024

YOU ARE NOT PERMITTED TO READ THE CONTENTS OF THIS QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY AN INVIGILATOR

ECN224 Econometrics 1 Duration: 2 hours

Answer ALL questions. The appendix contains tables with the critical values of the standard normal distribution and the F distribution.

Calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough workings in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

EXAM PAPERS MUST NOT BE REMOVED FROM THE EXAM ROOM

Please ensure that your working is clearly shown with all steps of your calculation included in your answer document, including any formula used.

Examiner: Dr. Štěpána Lazarová

The questions 1 and 2 below are based on a dataset containing data for full-time, full-year workers, ages 25–34, with a high school diploma or BA/BS as their highest degree. In particular, we will use the following variables:

- *ahe*: average hourly earnings in dollars.
- *female*: dummy variable equal to 1 if worker is a female, 0 otherwise.
- *bachelor*: dummy variable equal to 1 if worker has a bachelor's degree, 0 if worker has a high school degree.
- *age*: age of worker.

Question 1 (35 marks)

(a) A simple OLS regression of *ahe* on *age* has been carried out using homoskedasticity-only standard errors. The regression output is reported in Table 1. Interpret the estimated coefficient on *age*. Is this coefficient statistically significant? **(4 marks)**

(b) Is the sign of the coefficient on *age* in Table 1 as expected? **(4 marks)**

(c) Interpret the OLS estimate of the intercept in Table 1. Does the intercept have a real life meaning? **(4 marks)**

(d) What are two measures of fit reported in Table 1? Would you say this regression fits well? Why or why not? **(6 marks)**

(e) The regression of *ahe* on *age* has been re-estimated using heteroskedasticity-robust standard errors. The regression output is reported in Table 2. Provide the OLS estimates of the intercept and the coefficient on age in this regression. **(6 marks)**

(f) Using results in Table 2, formally test whether the coefficient is statistically significant at 1% and 5%. State your null and alternative hypothesis, give the value of your test statistic and its p-value, and conclude. Compare your answer to part (a). **(7 marks)**

(g) Would you say overall this regression model is a good model for estimating the effect of age on average hourly earnings? Why? **(4 marks)**

Table 1

```
> regr1 <- lm(ahe~age, data = cps)
> summary(regr1)

Call:
lm(formula = ahe ~ age, data = cps)

Residuals:
    Min      1Q  Median      3Q     Max 
-21.608  -8.176  -2.763   5.063  83.776 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.8826     1.4767   3.307 0.000949 ***
age           0.5520     0.0496  11.128  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.02 on 7096 degrees of freedom
Multiple R-squared:  0.01715,	Adjusted R-squared:  0.01701 
F-statistic: 123.8 on 1 and 7096 DF,  p-value: < 2.2e-16
```

Table 2

```
> coeftest(regr1, vcovHC(regr1, type = "HC1"))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) [        ]  1.439439   3.392 0.0006976 ***
age         [        ]  0.049102 [                  ]
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 2 (35 marks)

(a) The model in Question 1 has been enriched by regressing the logarithm of average hourly earnings, $\ln(ahe)$, on age, $female$ and $bachelor$ using heteroskedasticity-robust errors. The regression output is reported in Table 3. Interpret *all* of the estimated coefficients. **(4 marks)**

(b) Are the coefficients individually significant? **(4 marks)**

(c) Do the coefficients in Table 3 have expected signs? **(4 marks)**

(d) Considering the results in Table 3, are the three regressors in the regression in part (a) jointly significant? Justify your answer briefly.

(6 marks)

(e) The same regression as in part (a) has been estimated except that the binary variable $female$ was replaced by a binary variable $male$ (1 if a person is male, 0 if female). Fill in the blanks in the following equation, giving a brief explanation of how you obtained the values.

$$\ln\widehat{(ahe)} = \_\_\_\_\_\_\_\_ + \_\_\_\_\_\_\_\_ age + \_\_\_\_\_\_\_\_ male$$

$$+ \_\_\_\_\_\_\_\_ bachelor, \qquad R^2 = \_\_\_\_\_\_\_\_.$$

(5 marks)

(f) To explore possible nonlinearities in the relation between average hourly earnings and age, terms age^2 and age^3 have been included in the regression. The regression output is reported in Table 4. To decide whether this brings an improvement over the regression in part (a), a suitable test has been carried out. The results of this test are reported in Table 5. Which regression is the preferred one, the one in part (a) or part (f)? **(6 marks)**

(g) Considering all of the above results, does the analysis suggest that there is gender earning gap? That is, ceteris paribus, do females have lower average hourly earnings than men? Is the conclusion based on the reported results reliable? **(6 marks)**

Table 3

```
> regr2 <- lm(log(ahe)~age+female+bachelor, data = cps)
> summary(regr2)

Call:
lm(formula = log(ahe) ~ age + female + bachelor, data = cps)

Residual standard error: 0.4774 on 7094 degrees of freedom
Multiple R-squared:  0.2084,    Adjusted R-squared:  0.208
F-statistic: 622.4 on 3 and 7094 DF,  p-value: < 2.2e-16

> coeftest(regr2, vcovHC(regr2, type = "HC1"))

t test of coefficients:

              Estimate Std. Error t value  Pr(>|t|)
(Intercept)  2.0273590  0.0600122  33.782 < 2.2e-16 ***
age          0.0241912  0.0019966  12.116 < 2.2e-16 ***
female      -0.1776215  0.0115041 -15.440 < 2.2e-16 ***
bachelor     0.4615027  0.0114615  40.266 < 2.2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 4

```
> regr3 <- lm(log(ahe)~age+I(age^2)+I(age^3)+female+bachelor, dat
a = cps)
> summary(regr3)

Call:
lm(formula = log(ahe) ~ age + I(age^2) + I(age^3) + female +
    bachelor, data = cps)
```

$\boxed{}$

```
Residual standard error: 0.4773 on 7092 degrees of freedom
Multiple R-squared:  0.209,     Adjusted R-squared:  0.2085
F-statistic: 374.8 on 5 and 7092 DF,  p-value: < 2.2e-16

> coeftest(regr3, vcovHC(regr3, type = "HC1"))

t test of coefficients:

               Estimate  Std. Error  t value Pr(>|t|)
(Intercept) -1.09058321  8.27566297  -0.1318   0.8952
age          0.28920103  0.84826391   0.3409   0.7332
I(age^2)    -0.00714151  0.02883804  -0.2476   0.8044
I(age^3)     0.00005961  0.00032520   0.1833   0.8546
female      -0.17734220  0.01150084 -15.4199   <2e-16 ***
bachelor     0.46161317  0.01145403  40.3014   <2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table 5

```
> linearHypothesis(regr3, c("I(age^2)", "I(age^3)"), white.adjust
= "hc1")
Linear hypothesis test

Hypothesis:
I(age^2) = 0
I(age^3) = 0

Model 1: restricted model
Model 2: log(ahe) ~ age + I(age^2) + I(age^3) + female + bachelor

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F   Pr(>F)
1   7094
2   7092  2 2.9166 0.05418 .
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 3 (30 marks)

In a study of the relationship between real aggregate consumption and prices in a specific country based on a time series of 104 observations, the following linear model is considered,

$$C_t = \beta_0 + \beta_1 P_t + u_i. \tag{1}$$

The following OLS regression function was obtained (standard errors in parentheses),

$$C_t = \underset{(0.35)}{13.74} + \underset{(0.05)}{0.24} P_t, \qquad R^2 = 0.12, \qquad SER = 2.85, \tag{2}$$

where C_t is logarithm of consumption and P_t is logarithm of price.

(a) Is the coefficient on P_t statistically significant? **(4 marks)**

(b) Comment on the sign of the coefficient on P_t. Is it as expected?

(4 marks)

(c) Comment on the reliability of the OLS coefficient on P_t in equation (2). Is the OLS estimator unbiased? **(6 marks)**

(d) Propose an alternative method for estimating β_1. Explain when this estimator improves on the OLS estimator. **(6 marks)**

(e) Let Z_t be a productivity variable. The OLS regression of P_t on Z_t for $t = 1, \ldots, 104$ gives

$$\widehat{P}_t = -\underset{(0.45)}{3.01} - \underset{(0.07)}{0.74} Z_t, \qquad R^2 = 0.31, \qquad SER = 3.66.$$

Is Z_t a relevant instrument for price? **(6 marks)**

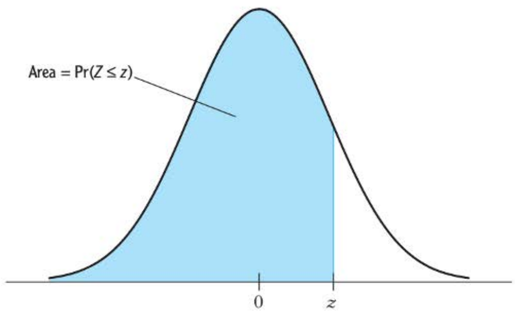
(f) The researcher computes an IV estimate of model (1) and obtains

$$\hat{\beta}_1^{IV} = -0.66, \quad \hat{\beta}_0^{IV} = 10.91.$$

Comment on the sign of $\hat{\beta}_1^{IV}$. **(4 marks)**

End of Paper - An Appendix of 3 pages follows

TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = \Pr(Z \leq z)$



z	Second Decimal Value of z: 0	1	2	3	4	5	6	7	8	9
–2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
–2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
–2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
–2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
–2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
–2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
–2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
–2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
–2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
–2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
–1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
–1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
–1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
–1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
–1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
–1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
–1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
–1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
–1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
–1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
–0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611

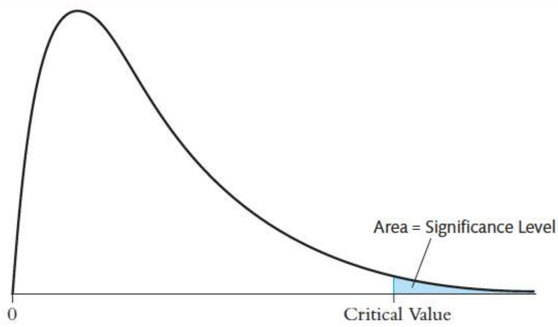
(Table 1 continued)

(Table 1 continued)

	Second Decimal Value of z									
z	0	1	2	3	4	5	6	7	8	9
–0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
–0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
–0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
–0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
–0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
–0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
–0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
–0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
–0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
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	0.7704	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

This table can be used to calculate $\Pr(Z \leq z)$ where Z is a standard normal variable. For example, when $z = 1.17$, this probability is 0.8790, which is the table entry for the row labeled 1.1 and the column labeled 7.

TABLE 4 Critical Values for the $F_{m,\infty}$ Distribution



	Significance Level		
Degrees of Freedom	**10%**	**5%**	**1%**
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64
8	1.67	1.94	2.51
9	1.63	1.88	2.41
10	1.60	1.83	2.32
11	1.57	1.79	2.25
12	1.55	1.75	2.18
13	1.52	1.72	2.13
14	1.50	1.69	2.08
15	1.49	1.67	2.04
16	1.47	1.64	2.00
17	1.46	1.62	1.97
18	1.44	1.60	1.93
19	1.43	1.59	1.90
20	1.42	1.57	1.88
21	1.41	1.56	1.85
22	1.40	1.54	1.83
23	1.39	1.53	1.81
24	1.38	1.52	1.79
25	1.38	1.51	1.77
26	1.37	1.50	1.76
27	1.36	1.49	1.74
28	1.35	1.48	1.72
29	1.35	1.47	1.71
30	1.34	1.46	1.70

This table contains the 90^{th}, 95^{th}, and 99^{th} percentiles of the $F_{m,\infty}$ distribution. These serve as critical values for tests with significance levels of 10%, 5%, and 1%.

End of Appendix