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

ECN322 Topics in Econometrics Duration: 2 hours

Answer all questions. Cross out any answers that you do not wish to be marked. There is a table of critical values from the standard normal distribution provided in the appendix.

You are permitted to bring 20 x A4 pages of notes into your examination (i.e. 10 double sided pieces of paper). These can be typed or handwritten and can include graphs and images. They can include material from any source.

Your notes must be stapled together and include your student ID number and the module code on the first page. You must submit your notes at the end of the examination with your answer booklet.

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

Question 1 (34 marks)

(a) Use the fundamental probability transform and a generator of standard uniform $U(0,1)$ observations to draw a sample of $n = 800$ observations from the standard logistic distribution. The cumulative distribution function of the standard logistic distribution is

$$F(x) = \frac{1}{1 + e^{-x}}.$$

(6 marks)

(b) Using generators of random draws from the standard uniform $U(0,1)$ and standard normal $N(0,1)$ distributions, produce 500 draws from the following three distributions:

(i) flips of a fair coin,

(ii) $N\left(1, \frac{1}{4}\right)$,

(iii) $F(4,3)$. **(4 marks)**

(c) Consider a linear model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

where $E(\varepsilon_i|X_i) = 0$, $\operatorname{var}(\varepsilon_i|X_i) = \sigma^2$ and where errors ε_i are uncorrelated. Select specific values of β_0, β_1 and σ^2 and draw a sample of $n = 250$ draws of (Y_i, X_i) from this model. Plot a scatterplot of Y_i against X_i. **(6 marks)**

(d) Now run a regression of Y_i on X_i and intercept using data generated in part (c). Are the estimated parameters from this regression close to the values you used when generating the data? Comment on the results.

(3 marks)

(e) Carry out a Monte Carlo study of the distribution of the sample variance of a random sample of $n = 20$ observations from the $t(6)$ distribution. Generate a sample $x_1, \ldots, x_n$ with $n = 20$ from the $t(6)$ distribution as follows. Compute the sample variance (EViews function `@vars`) and store its value. Repeat $10,000$ times. Plot a histogram of the resulting $10,000$ values of the sample variance. Is the distribution of the sample variance a normal distribution? If not, how does this distribution differ from the normal distribution? **(6 marks)**

(f) Repeat the experiment from part (e) for a sample $x_1, \ldots, x_n$ with $n = 100$. Is the distribution of sample variance approximately normal in this case? How does your answer compare to the answer you gave in part (e)? **(3 marks)**

(g) The central limit theorem (CLT) implies that for large n, the sample variance of a sample of n i.i.d. variables with finite kurtosis is approximately normally distributed. Is sample size $n = 30$ or $n = 100$ sufficiently large for the CLT approximation to be satisfactory? If not, for which sample size would the approximation be satisfactory? Illustrate your opinion with appropriate simulation output. **(6 marks)**

Turn Over

Question 2 **(33 marks)**

(a) Suppose you are given a random sample $x_1, \ldots, x_{143}$ from a normal distribution. Provide a simple example of a parametric and nonparametric model of probability density of the underlying distribution of this data. For each model, describe the properties or features of the probability density function. **(4 marks)**

(b) Give at least two advantages and two disadvantages of nonparametric estimation when compared to parametric estimation. **(4 marks)**

(c) Define a kernel regression estimator. Describe the role of the components of the estimator. **(5 marks)**

(d) When using a kernel density or kernel regression estimator, a kernel and a bandwidth need to be selected. Which of the two choices is more important and why? **(4 marks)**

(e) Describe the variance-bias trade-off occurring when comparing different values of bandwidth in the kernel regression estimator. Given this trade-off, what is the optimal choice of bandwidth? **(4 marks)**

(f) Suppose you have a sample of values which you know come from a $\chi^2(k)$ distribution but you do not know the degrees of freedom parameter k. If you wish to estimate the density of the data, would you recommend a parametric or nonparametric estimator? Explain your recommendation. **(6 marks)**

(g) Let the nonparametric regression model be given as

$$y_i = m(x) + \varepsilon_i, \qquad i = 1, \ldots, n.$$

Given a sample $(x_1, y_1), \ldots, (x_n, y_n)$, an estimate of the function m can be obtained be minimizing a weighted sum of squares

$$S(x) = \sum_{i=1}^{n} (y_i - m(x))^2 K\left(\frac{x - x_i}{h}\right)$$

with respect to $m(x)$ where weights are given by $K\left(\frac{x-x_i}{h}\right)$ with K a kernel and h a bandwidth. Derive a local mean estimator of the regression function g by minimizing $S(x)$. **(6 marks)**

Question 3 (33 marks)

(a) Let U be a standard uniform random variable, $U \sim U(0,1)$. Let Y be a random variable with the following distribution:

$$Y = \begin{cases} 1 & \text{with probability } 1/4, \\ U+1 & \text{with probability } 3/4. \end{cases}$$

Write down the cumulative distribution function of Y and sketch a plot of this function. **(4 marks)**

(b) Sketch a plot of the quantile function for the random variable Y in part (a). **(4 marks)**

(c) A sample of data on yearly income ($YInc$) in £1,000s for employees with varying years of education (Edu) has been collected. For employees with 12 years of education, the conditional quantiles of yearly income are as follows:

τ	10%	25%	50%	75%	90%
$Q_{YInc}(\tau \mid Edu = 12)$	18.2	26.3	31.0	43.5	83.3

Compute the corresponding conditional quantiles of monthly income ($MInc$) for employees with 12 years of education. **(4 marks)**

(d) In the sample in part (c), yearly income has been subsequently topcoded and incomes of over £50,000 were recorded as £50,000. How do the conditional quantiles of yearly income in part (c) change for employees with 12 years of education? **(4 marks)**

(e) Let X be a random variable. Show that the median of X minimizes $E|X - \theta|$, that is, that

$$\text{med}(X) = \arg\min_{\theta \in R} E|X - \theta|.$$

Assume that X has a continuous distribution. **(6 marks)**

Question continues on next page...

(f) A study of the effect of various factors on housing prices (in millions of new Taiwan dollars) in Taiwan has been carried out using a dataset of 76,264 observations. Linear mean regression and linear quantile regression of the logarithm of the property price on the factors such as location or age of the property were run. The coefficients on *Age* are reported in Table 1.

Which of the coefficients are significant at the 5% significance level? Justify your conclusion. **(3 marks)**

Table 1: Estimated coefficients on the age of the property. The numbers in parentheses are the t-values of the coefficients.

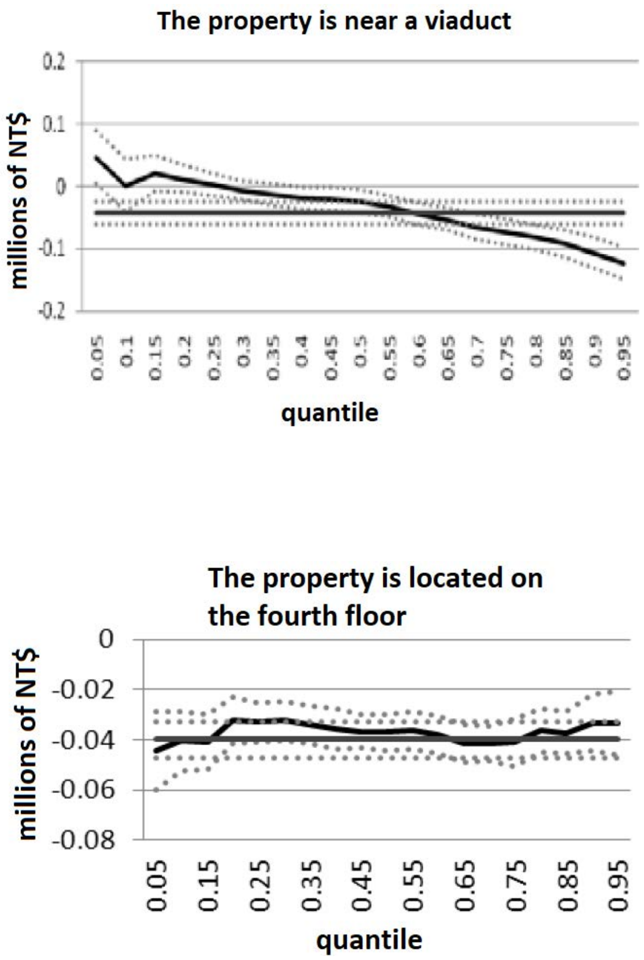
	OLS estimate	Quantile Regression Estimates				
	Coeff.	0.1	0.25	0.50	0.75	0.9
Age	-0.006 (-34.29)	-0.012 (-31.58)	-0.003 (-37.48)	-0.01 (-37.66)	-0.004 (-18.23)	-0.002 (-6.30)

(g) Figure 1 below reports estimated quantile regression coefficients for two indicator variables characterizing the location of the property. The solid dark grey straight line is the OLS fitted line, the solid black line depicts the quantile estimates, and the dotted lines describe 95% confidence bands for the coefficient estimates.

In both cases, interpret the coefficient estimates. Does quantile regression bring additional information compared to the OLS regression? Are the effects of the two location variables different for different quantile levels? Do the location variables appear to exert a pure location shift effect on the conditional distribution of prices? **(8 marks)**

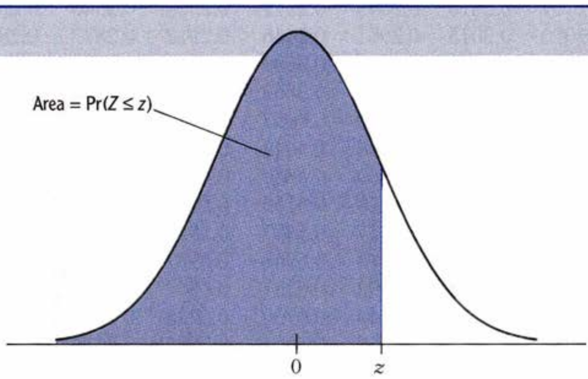
Question continues on next page...

Figure 1: Estimated quantile coefficient for two indicator variables describing the location of the property



End of Paper - An Appendix of 2 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

continued on next page

TABLE 1 *(continued)*

z	Second Decimal Value of 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.

End of Appendix