School of Economics
ECO-7000A: Econometric Methods

ONLINE EXAMINATION

To be taken on Tuesday 16 January 2024

- This examination paper will be available for **FIVE HOURS**, from 9.30 AM UK time. The submission point will close at 2.30 PM UK time, unless you are entitled to Assessment Adjustments.

- It is recommended yohu spend a maximum of **4 hours** completing this examination and **1 hour** checking and uploading your work.

- Answer ALL THREE questions.

- Questions are divided into parts. There are 25 parts in total. All parts carry equal marks (i.e. 4% each).

- Write (or type) your answers in the spaces provided.

- Some parts simply require you to insert numerical answers in tables. Other parts require you to provide interpretation and discussion of results. In the latter case, you are advised not to exceed **100 words** in your answer to any single part.

- Written material MAY be consulted.

To maintain academic integrity students are required:

- Not to keep an electronic copy of this paper or share the content with others

- To ensure that all work produced is solely their own. Students are advised that their work will be checked to determine any similarity with the work of other students.

The three data sets required to answer the two questions in this test are:

- **SAVINGS_2023**
- **BIGMAC_ONEYEAR**
- **VOTING**

WRITE YOUR STUDENT NUMBER BELOW:

STUDENT NUMBER: __100455844_____

QUESTION 1 (40 marks)

The file **SAVINGS_2023** contains data from a developed economy (**quarterly; 1994:1 q1 through to 2023:q3**) on the following variables:

t:	Time trend
q:	Quarter (1 if JFM; 2 if AMJ; 3 if JAS; 4 if OND)
s:	Real savings (£billion at 2023 prices)
y:	Real personal disposable income (£billion at 2023 prices)
r1:	Short term interest rate (%)
r2:	Long term interest rate (%)

(a) What were short term and long term interest rates in the second quarter of 1997?

R1 = 3.25
R2 = 3.6

(b) Estimate a regression model with **s** as the dependent variable, and **y** as the explanatory variable. Report the coefficients and standard errors in the following table (**USE 4 DECIMAL PLACES**). Also report the R^2 and the sample size.

Variable	***Coefficient***	***Standard error***
y	0.3756	0.0319
Constant	117.2604	12.1867
R-squared	0.5418	
Sample size	119	

(c) What is your estimate of the marginal propensity to save (MPS)? **[Note: MPS is just the coefficient of y in the regression.]** Report the 95% confidence interval for MPS. Interpret this interval.

Estimate of MPS: 0.3756

95% confidence interval for MPS: The 95% confidence interval for this MPS is between 0.3124 and 0.4389. This range suggests that we can be 95% confident that the actual MPS value lies within these bounds. The fact that zero is not included within this interval signifies that the relationship between income and savings is statistically significant – indicating that an increase in income is positively correlated with an increase in savings. The relatively narrow span of this interval points to a considerable level of precision in this estimate, reinforcing the likelihood that the observed relationship is not a product of random variations.

(d) Test for first-order and fourth-order serial correlation in the model of (b) using the Breusch-Godfrey test. Report the results in the table below.

Lags	***Chi2***	***df***	***p-value***
1	**14.444**	**1**	**0.0001**
4	**18.735**	**4**	**0.0009**

What does this tell us about the model of (b)?

Conclusion: Since the first p-value is less than 0.01, there is STRONG EVIDENCE of first order serial correlation. Since the second p-value is less than 0.05, there is EVIDENCE of fourth order serial correlation.

This tells us that there is a problem in the model of (b), and this problem is probably caused by missing variables.

Presence of Serial Correlation: The significant results for both first-order and fourth-order serial correlation suggest that the model may have omitted variables that are influencing s (real savings), or there may be underlying time-dependent processes not captured by the model. The residuals are not independent across time, which is an assumption of standard linear regression models

(e) Add the two interest rates (r1 and r2) to the regress command used in (b). DO NOT REPORT THE RESULTS. Test for first and fourth order serial correlation again.

Lags	*Chi2*	*df*	*p-value*
1	**0.099**	**1**	**0.7534**
4	**7.604**	**4**	**0.1072**

Explain the difference between the results of (d) and (e).

The addition of **r1** and **r2** appears to have improved the model by accounting for factors that were previously omitted. These interest rates might be capturing economic conditions or influences that affect savings behavior, and their inclusion helps explain the variation in savings more effectively.

The significant change in the Breusch-Godfrey test results suggests that short-term and long-term interest rates are important variables in explaining savings. They might be capturing the effects or trends that were previously manifesting as serial correlation in the residuals.

(f) Are either of the two interest rates (r1 and r2) important in explaining savings behaviour, according to tests of individual significance? First, report the t-statistics and p-values. Then draw conclusions.

Short-term interest rate: $t_{r1} = 1.22$ $p_{r1} = 0.223$

Long-term interest rate: $t_{r2} = 1.74$ $p_{r2} = 0.085$

Conclusions: The p-value for the long-term interest rate is 0.085, which is also above the conventional 0.05 threshold but lower than that of the short-term interest rate. This indicates that while the long-term interest rate has a stronger association with savings behavior than the short-term rate, it still does not reach the level of statistical significance typically used to confirm such a relationship.

(g) Are the two interest rates jointly significant according to an F-Test? **You may use the `test` command in STATA**. Quote the F-statistic and the p-value.

F-Statistic: 99.98
p-Value: 0.0000

The F-statistic of 99.98 is substantially high, and the corresponding p-value is 0.0000, which is very close to zero. This indicates a strong rejection of the

null hypothesis that both **r1** (short-term interest rate) and **r2** (long-term interest rate) have no joint effect on the dependent variable **s** (real savings). The extremely low p-value strongly suggests that the two interest rates, when considered together, are jointly significant in explaining the variation in savings behavior. This means that, as a pair, the short-term and long-term interest rates provide important information about savings behavior in your model.

(h) By computing a correlation matrix of the explanatory variables, investigate the extent to which multicollinearity is causing problems in this model. Insert your correlations in the correlation matrix below.

	y	r1	r2
y	1		
r1	-0.0945	1	
r2	-0.0736	0.9780	1

Could multicollinearity be an explanation for any apparent contradiction between your answers to (f) and (g)?

In the context of our regression analysis (parts (f) and (g)), the high degree of multicollinearity between the interest rates might explain why they appeared not individually significant in part (f) but were jointly significant in part (g). When variables are highly correlated, it can be challenging to isolate their individual effects on the dependent variable, as they may be capturing similar underlying economic phenomena.

(i) Add `i.q` to the `regress` command used in (e), in order to include a set of quarterly dummy variables. DO NOT REPORT THE REGRESSION RESULTS. Conduct an F-test of the significance of season in explaining savings behaviour. **Use the F-test formula on the formula sheet (available on the module Blackboard page), and verify that your answer is correct using the `test` command in STATA.** Interpret the result.

$R_u^2 = ¿$

$R_r^2 = ¿$

$$F = \frac{\left(R\text{U}^2 - R^2\text{R}\right)/\text{r}}{(1 - R\text{U}^2)/(n - k)}$$

$$F = \frac{(0.8327 - 0.5418)/2}{((1 - 0.8327)/(119 - 4)}$$

$$F = 99.98$$

$$F_{2,115,0.05} = 3.08$$

(j) Report the coefficients of the quarterly dummies in the model of (i). Interpret them by describing the seasonal pattern in savings behaviour. In which quarter is savings highest, and in which quarter is it lowest? Can you explain this seasonal pattern?

Thank you for providing the coefficients for each quarter, with Quarter 1 (JFM) as the base. Here's the corrected interpretation:

Coefficient: -2.977
- Savings in Quarter 2 (April to June) are significantly lower compared to Quarter 1 (JFM). This suggests that people tend to save less during the spring months, possibly due to increased spending on springtime activities and outdoor leisure.

- Coefficient: -4.404
- The coefficient for Quarter 3 (July to September) is even more negative, indicating significantly lower savings compared to Quarter 1. Summer vacations and increased spending on leisure activities during the summer months may contribute to reduced savings during this period.

- Coefficient: -3.258
- Savings in Quarter 4 (October to December) are also significantly lower compared to Quarter 1. This suggests a seasonal pattern where savings tend to decline during the holiday season, possibly due to increased spending on gifts and festivities.

In reality quarter 1 savings is is the lowest, considering that most people are probably just recovering from spending that occurred during quarter 4 of the year.

END OF QUESTION 1

QUESTION 2 (20 MARKS)

The file **BIGMAC_ONEYEAR** contains data on 54 countries in July 2023. The variables that you will need to use are:

p_local: price of a Big Mac (the McDonald's hamburger) in local currency in July 2023

p_usa: price of Big Mac in the USA in July 2023 (note: this is the same value in every row)

e: exchange rate for against the US dollar in July 2023 (that is, *e* is the number of units of local currency that can be exchanged for one US dollar in July 2023).

(a) From the available data, compute the exchange rate for the Indonesian Rupiah against the Qatari Rial 2023 (i.e. find the number of Indonesian Rupiah that can be exchanged for one Qatari Rial).

15080 Indonesia rupiah= 3.64 Qatar rial (both = 1 US$)

$\Rightarrow$ 1 Qatari rial = 15080 /13 = 4142.86 Indonesia rupiah

(b) By dividing p_local by e, compute the price of a Big Mac in each country <u>in US dollars</u> (do NOT report the complete results). On this basis, which of the 54 currencies appears the most under-valued in 2023, and which the most over-valued?

Taiwan TWD $2.39 most under-valued currency Swiss Franc $7.73 most over-valued currency

Generate the required variables, and estimate the following regression model:

$$\log\left(\frac{p\_local_i}{p\_usa}\right) = \beta_1 + \beta_2 \log e_i + u_i \; ; \; i = 1, \cdots, 54 \qquad (1)$$

Do not report the results.

(c) In Model (1), perform a test of the joint null hypothesis $H_0 : \beta_1 = 0; \beta_2 = 1$. Interpret the result of the test. Which theory is being tested?

```
. test (log_e=1) (_cons = 0)
```

```
( 1)  log_e = 1
( 2)  _cons = 0

      F(  2,    52) =   29.10
           Prob > F =    0.0000
```

This is a test of the Law of One Price (LOP). LOP implies that the price of a bigmac is the same in all countries. Since the p-value is less than 0.01, we are strongly rejecting LOP.

(d) Again in Model (1), perform a test of the simple null hypothesis $H_0 : \beta_2 = 1$. Interpret the result of the test. Which theory is being tested?

```
. test (log_e=1)

 ( 1)  log_e = 1

       F(  1,    52) =    2.47
            Prob > F =    0.1223
```

This is a test of Purchasing Power Parity (PPP). PPP implies that prices in the two countries adjust to accommodate changes in exchange rates. Since the p-value is greater than 0.01, we are accepting PPP.

(e) Using the Breusch-Pagan test, test for heteroscedasticity in Model (1). Report the test statistic and the p-value. Does the result indicate that the tests performed in (c) and (d) are valid?

. hettest

Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of log_p_ratio

H0: Constant variance

chi2(1) = 0.79
Prob > chi2 = 0.3740

END OF QUESTION 2

QUESTION 3 (40 MARKS)

The file **VOTING** contains UK data on voting behaviour based on the British Election Survey (note that Scotland is excluded from the data). The available variables are:

voted:	=1 if respondent voted in last local election; = 0 if did not vote
age:	Age in years
female:	=1 if female; =0 if male
region:	=1 if respondent lives in North, =2 if in Midlands or Wales, =3 if in South or London
LeftRight:	Left-Right political spectrum (0 = far Left-wing - 10 = far Right-wing)
degree:	=1 if respondents has acquired a higher education degree, =0 otherwise
party:	Which party do you feel closest to? : =1 if Conservatives =2 if Labour =3 if Labour Democrats

(a) How many respondents are in the survey and what proportion of them voted in the last local election? Also, out of those who voted, what proportion feels closest to the Labour party?

Number of respondents in survey: 640

Proportion voted: 86.09%

Out of voters only, proportion feeling closest to the Labour party: 35.21%

(b) Estimate a Probit model, with **voted** as the dependent variable, and with **age**, **female, degree** and **LeftRight** as explanatory variables. Report coefficients and standard errors in the following table (**USE 4 DECIMAL PLACES**). Also report the (maximised) log-likelihood, and the sample size.

Variable	***Coefficient***	***Standard error***
age	0.0169	0.0039
female	0.2294	0.1283
degree	0.4617	0.1397
LeftRight	0.0501	0.0324
constant	-0.0381	0.2874
LogL	-242.2664	
Sample size	640	

(c) Using your Probit results in (b), obtain the *average marginal effect* (AME) for variable **female**. Report and interpret this AME, and also comment on its statistical significance.

AME for **female**: 0.0477

Interpretation: This means that, on average, being female (as opposed to male) increases the probability of voting in the last local election by about 4.77 percentage points, holding all other variables constant.

Statistical Significance:

The positive sign of the AME, which is 0.0477, indicates a trend where females are more likely to vote compared to males. However, given the lack of statistical significance at the 5% level, as indicated by the p-value of 0.073 and a z-statistic of 1.79, this result should be interpreted with caution. It suggests that while there may be a tendency for females to vote more than males, the evidence, with a standard error of 0.0266 and a confidence interval ranging from -0.0045 to 0.1000, is not strong enough to conclusively assert this difference in the population from which this sample was drawn.
The borderline p-value, which is close to 0.05, along with the z-statistic of 1.79 and the confidence interval that barely includes zero, suggest that while the effect is not statistically significant at the 5% level, it is at a level close to significance. This could warrant further investigation or could be indicative of a small but existent effect in the population.

(d) Using your Probit results in (b), calculate the predicted probability of having voted in the last local election for two hypothetical individuals with the following characteristics:

(i) a 92-year-old female, with a degree and with a LeftRight score of 10
(ii) an 18-year-old male, without a degree and with LeftRight score of 1

At age 92 female with a degree and with a LeftRight score of 10 = 99.09%

- At 18-year-old male, without a degree and with LeftRight score of 1 = 46.93%

(e) Obtain the Brier Score for the model estimated in (b). What does this measure tell us?

Brier Score = 0.1135
Since we want Brier Score that are smaller than 0.25 and the closer it is to zero the better the performance of the model. Here with a score of 0.1135 the model is strong because of it's proximity to 0.

(f) For the model estimated in (b), obtain the sensitivity, specificity and the overall % of correct classification setting the threshold probability at 0.5. Interpret your results and explain what these results suggest about the predictive performance of the model.

Sensitivity = 100.00%

Specificity = 0.00%

% of correct classification = 86.09%

Sensitivity

- **Sensitivity (100.00%):** This means that the model correctly identified 100% of the actual cases where individuals voted (True Positives). In other words, there were no voters that the model incorrectly classified as non-voters. While this seems ideal, a sensitivity of 100% can sometimes indicate that the model is too lenient in predicting the positive class (voted).

Specificity

- **Specificity (0.00%):** The model failed to correctly identify any of the actual non-voter cases (True Negatives). A specificity of 0% means that all non-voters were incorrectly classified as voters. This is a significant issue as it indicates that the model is not effective at distinguishing those who did not vote.

Overall Accuracy

Correctly Classified (86.09%): Despite the high overall accuracy, this metric is misleading in this context. The model appears to perform well in terms of overall accuracy solely because it classified nearly everyone as a voter, and the majority of the dataset consists of voters.

- **Over-Prediction of Voting:** The model is over-predicting the voting behavior, classifying almost everyone as a voter. This suggests that the model might be biased towards predicting the majority class in the dataset.
- **Lack of Discrimination:** The model lacks the ability to discriminate between voters and non-voters. A good predictive model should have a balance between sensitivity and specificity.

- **Model Assessment:** These results suggest that the model may not be suitable for making reliable individual-level predictions about voting behavior, especially in identifying non-voters.
- **Model Adjustment Needed:** Consider revisiting the model, including feature selection, threshold adjustment, or trying different modeling techniques to improve its ability to distinguish between voters and non-voters.

(g) Add the **region** variable the the Probit model used in part (b). Make sure you use dummy variables for each region, keeping 'Midlands or Wales' as the base group. Report coefficients and standard errors in the following table (**USE 4 DECIMAL PLACES**). Also report the (maximised) log-likelihood, and the sample size.

Variable	***Coefficient***	***Standard error***
age	0.0168	0.0039
female	0.2295	0.1283
degree	0.4665	0.1401
LeftRight	0.0533	0.0331
North	0.1010	0.1576
South or London	0.0905	0.1587
constant	-0.4597	0.3090
LogL	-242.0198	
Sample size	640	

(h) Conduct a Likelihood Ratio (LR) test of the significance of region of residence in predicting whether an individual voted or not in the local elections. Report the LR test statistic, the p-value of this test, and provide an appropriate conclusion based on this test.

LR test stat:0.49

p-value:0.7814

Interpretation: Since the p-value is above 0.01, there is no substantial evidence to suggest that the region plays a significant role in determining whether a person voted in the local election.

(i) An additional variable in the data set is named **party**, representing which political party the respondent feels closest to. It is defined as follows:

=1 if feeling closest to Conservatives party
=2 if feeling closest to Labour party
=3 if feeling closest to Liberal Democrats party

Estimate a Multinomial Logit model (command mlogit in Stata) with **party** as the dependent variable, using the same explanatory variables as in part (g). Report your Stata mlogit command below and briefly summarise your key findings. You don't need to present the estimated results.

Command: mlogit party age female degree LeftRight ib2.region

Model Overview

- **Dataset Size:** The analysis was performed on 640 observations.
- **Model Fit:** The log likelihood at convergence is -443.48535, which, along with the LR chi2 of 342.08 and a highly significant Prob > chi2 (0.0000), suggests that the model fits the data reasonably well.
- **Pseudo R2:** The Pseudo R2 value of 0.2783 indicates a moderate relationship between the predictors and the response variable.

Party Affiliation as Outcome

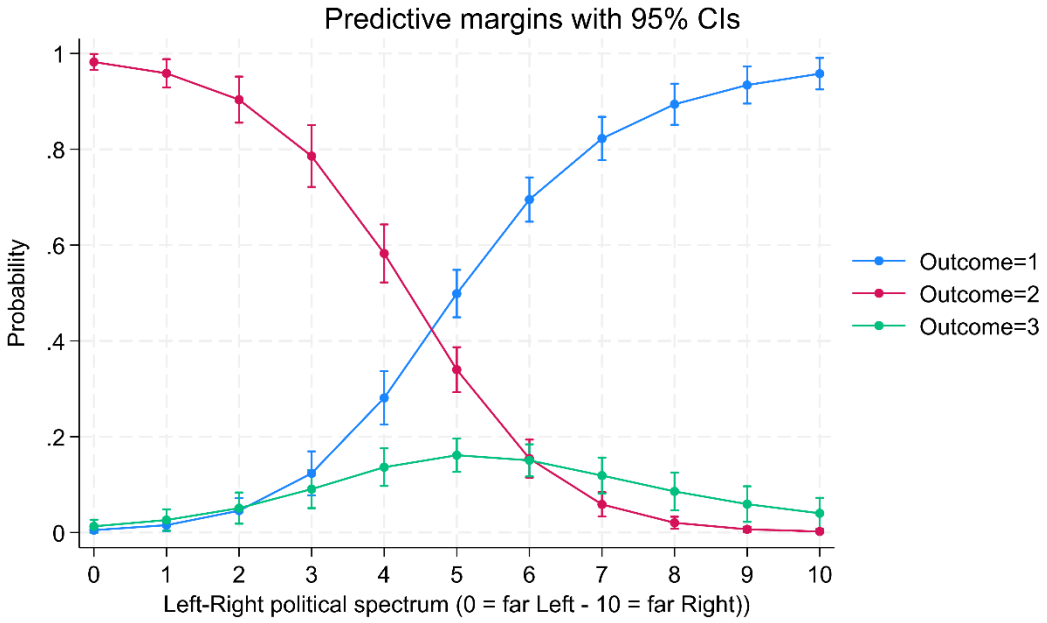
- The model predicts the probability of an individual feeling closest to the 'Labour' or 'Liberal Democrats' parties, compared to the base outcome of 'Conservatives'.

Key Findings for Each Party

- **Labour:**
  - **Age:** The negative coefficient for age suggests a decreasing likelihood of feeling closest to Labour with increasing age, though this effect is not statistically significant ($p = 0.239$).
  - **Gender:** Being female (positive coefficient) seems to slightly increase the likelihood of preferring Labour, but this is not statistically significant ($p = 0.562$).
  - **Education (Degree):** Holding a degree is negatively associated with preferring Labour, but again, this is not statistically significant ($p = 0.217$).
  - **LeftRight:** A strong negative coefficient indicates that individuals with higher Left-Right scores (more right-wing) are significantly less likely to prefer Labour ($p < 0.001$).
  - **Region:** Living in the North significantly increases the likelihood of preferring Labour compared to the Midlands or Wales, while living in South or London shows a negative but not significant association.
- **Liberal Democrats:**
  - **Age:** Similar to Labour, a negative coefficient for age, but not statistically significant ($p = 0.536$).
  - **Gender:** Being female decreases the likelihood of preferring Liberal Democrats, though not significantly ($p = 0.427$).
  - **Education (Degree):** Having a degree significantly reduces the likelihood of preferring Liberal Democrats ($p = 0.003$).
  - **LeftRight:** A negative coefficient suggests that more right-leaning individuals are less likely to prefer Liberal Democrats, with statistical significance ($p < 0.001$).

- **Region:** Residing in the North or in South or London does not significantly influence the preference for Liberal Democrats compared to Midlands or Wales.

(j) Using the results of the model in (i), provide a graph of predicted probabilities for each of the outcomes of the **party** variable, and for **LeftRight** values between 0 and 10 (with increments of 1). Make sure you report the Stata commands used to obtain this plot. Based on this graph, explain how the probability of voting for each of the parties depends on the **LeftRight** political spectrum.



Interpretation:

- The graph suggests that political alignment (Left-Right orientation) is a strong predictor of voting behavior for the Conservative and Labour parties, which show clear and opposing trends in relation to the political spectrum.
- The Conservative party (Outcome=1) has increasing support as individuals' political views become more right-wing.
- The Labour party (Outcome=2) has strong support on the Left which diminishes substantially as the spectrum moves towards the Right.
- The third party's support (Outcome=3) does not appear to be strongly influenced by the Left-Right spectrum, indicating that other factors may be more relevant for predicting their voter base.
- The confidence intervals for each predictive margin suggest that there is some uncertainty around these estimates, but the trends are clear for the Conservative and Labour parties. The confidence intervals are wider for the third party, indicating more uncertainty about the predictive margins for this group across the political spectrum.

The graph overall highlights the significant role of the Left-Right political spectrum in shaping the probability of voting for major parties, with a clear distinction between the preferences of left-wing and right-wing individuals.

END OF TEST