ECON 4403 ECONOMETRICS

Final Assignment

SMU - Dr. Aydede

Winter 2021

It's due on **April 19 by 5pm**. You can submit it individually or as a team (max 3) by using Dropbox on Brightspace.

- You can use R, Stata, Stata – Mata.
- What you submit must be your own work.
- Pay attention to our format restrictions as stated in A1. I need ONLY **rmd and html** files (or **smcl** if you use Stata). Please report the result after each operation **ONLY if they are not big matrix or dataframe**. In that case, you can show the `dim()` and `head()`.

Q1 (You can only use linear algebra in this question - Each 4 points)

The study of wage earning determination has a long history. According to one theory, the human capital theory, the wage earning reflects the labour market rewards of human capital, a set of skills that a person owns. The determination of the wage earning is usually estimated with log(wage) in the following based model:

$$\ln wage = \beta_1 + \beta_2 educ + \beta_3 exper + \beta_4 tenure + \text{e}$$

Here, `wage`= Average hourly wage earning, `educ` = years of education, `exper`= years of potential experience, and `tenure` = years with the current employer.

1. Estimate the model by using `wage.dta` and interpret the results. Calculate R^2 and $\bar{R}^2$.
2. Do you think the `ln(wage)` is linearly related to `exper` and `tenure`? Explain it by estimating the adjusted base model given above.
3. According to the human capital theory, gender, skin color, and marital status are not a part of skill sets. Expand your model in (1) to include `nonwhite`, `female`, and `married`. What do you think about the results?
4. Use a F-test to decide if the expanded model in (3) is accepted.
5. Test for heteroskedasticity. If you have the problem, calculate the correct standard errors.
6. Test for serial correlation by using DW test.
7. Suppose the true model is $\ln wage = \beta_1 + \beta_2 educ + \beta_3 exper + e$. But you omit `exper` and estimate an **underfitted** model: $\ln wage = \alpha_1 + \alpha_2 educ + v$. Show that the bias is:

$$\hat{\alpha}_2 - \beta_2 = \beta_3[\mathbf{cov}(educ, exper)/\mathbf{var}(educ)]$$

8. Suppose the true model is $\ln wage = \beta_1 + \beta_2 educ + \beta_3 exper + e$. But you add an irrelevant variable, `xexper` (`xexper <- rnorm(526, 50, 2)`) and estimate an **overfitted** model: $\ln wage = \beta_1 + \beta_2 educ + \beta_3 exper + \beta_4 xexper + e$. Show that the overfittted model inflates the stadard errors of estimators.

Q2. (Each 4 points)

The capital asset pricing model (CAPM) is well known in finance. It explains variations in the rate of return on a security as a function of the rate of return on a portfolio consisting of all publicly traded stocks (the so-called market portfolio). Generally the rate of return on an investment is measured relative to its opportunity cost, which is the risk-free return. The resulting difference is called the risk premium, since it is the reward or punishment for making a risky investment. The CAPM says that the risk premium on a security is proportional to the risk premium on the market portfolio. That is,

$$r - r_f = \beta \left(r_m - r_f \right)$$

where r = return to a security, r_f = risk-free return, and r_m = return on the market portfolio, and β is that security's "**beta**" value. A stock's beta is important to investors since it reveals the stock's volatility. It measures the sensitivity of that security's return to variation in the whole stock market. As such, values of beta less than 1 indicate that the stock is "defensive" since its variation is less than the market's. A beta greater than 1 indicates an "aggressive stock". Investors usually want an estimate of a stock's beta before purchasing it. Consider the following model for examining this issue:

$$rpr = \gamma + \beta(mpr) + e$$

where $rpr = r – r_f$ = risk premium on a security, $mpr = (r_m – r_f)$ = risk on the market portfolio.

Answer the following questions using the data file `capm.dta` on the monthly returns of six firms – Disney (`dis`), GE (`ge`), GM (`gm`), IBM (`ibm`), Microsoft (`msft`), and Mobil-Exxon (`xom`), the rate of return on the market portfolio (`mkt`), and the rate of return on the risk-free asset (`rf`). The 132 observations cover January 1998 to December 2008.

1. Estimate the model for each firm and present your results. You will have a total of six regressions, one for each firm. You would need to generate the dependent and independent variable for each firm.
2. Comment on the estimated β values. Which firms appear aggressive? Defensive? (You can write a short text on your do file that prints on your smlc file)
3. Test whether these β values are significantly different from 1 by using F-test.
4. Finance theory says that $\gamma = 0$. Does it seem to be confirmed by your results? (Comment 1-2 sentences)

Q3. (Each 4 points).

The study of the link between inflation and the supply of money has a long history. According to one theory, inflation is closely tied to the rate of money growth as well as the rate at which output grows. Consider the following equation that models this relationship:

$$\text{INF} = \beta_1 + \beta_2 \text{GM} + \beta_3 \text{GX} + \text{e}$$

Here, INF= rate of inflation (%), GM= rate of growth of the money supply (%), and GX= rate of growth real GDP (%).

The file `money.dta` contains data on these variables for a sample of 76 countries and is taken from a study by Harold J. Brumm (2005), Money Growth, Output Growth, and Inflation: A Reexamination of the Modern Quantity Theory's Linchpin Prediction.

1. Plot `INF` against `GM` and `GX` separately.
2. According to the quantity theory, $\beta_2 = 1$, and $\beta_3 = -1$. Conduct F-test to assess whether the evidence supports these restrictions. Comment on it.

3. How would you test the restriction $\beta_2 + \beta_3 = 0$ by rewriting the model. Estimate the new model and present your main results. Then test whether this restriction is supported at the 5% level of significance. Comment on it.

4. Test for autocorrelation by using DW. Apply the first-difference GLS if the DW test indicates $\rho = 1$.

Q4. (Each 4 points)

This question will apply a dummy variable model by using our `wageedu.xlsx` data.

1. Find conditional means $\mathbf{E}(Y|X)$.
2. Create four dummy variable: D1 = 1 if schooling = 6 otherwise 0, D2 = 1 if schooling = 8 otherwise 0, D3 = 1 if schooling = 12 otherwise 0, D4 = 1 if schooling = 16 otherwise 0. Calculate $\mathbf{E}(Wage|Schooling)$ by using a dummy variable model without a constant. Are they the same as in (1)?
3. Is the average wage for those who have 12 years of schooling **significantly** different than that for those with 16 years of schooling?

Q5. (Each 4 points)

Use `utown` data and create an indicator variable from `price` that reveals if the house in the top 25 percentile. Call it `High`, which will be 1 if the house price higher than (and equal to) the threshold, 0 otherwise. Here is the model that we want to estimate:

$$\text{High}_{\;i} = \beta_1 + \beta_2 \text{age}_{2i} + \beta_3 \text{sqft}_{3i} + \beta_4 \text{ pool }_{4i} + \beta_5 \text{ fplace }_{5i} + u_t$$

1. Estimate linear probability model (LPM) and interpret the results.
2. Estimate logistic regression and Interpret the coefficients.
3. Calculate and interpret odd ratios (OR).
4. Calculate the marginal effects at means. Are they different than the ones given by LPM?
5. Predict the probabilities.
6. Based on your predictions, identify the models success. What's the "false positive" rate?