TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

Mock Exam

Quantitative Methods III

1. Consider the following estimated models on a sample of 5,000 workers. (Dependent variable: AHE, Average Hourly Earnings in dollars):

Regressor	(1)	(2)
Intercept	12.205 (0.578)	13.647 (0.660)
Age (X_1)	0.203 (0.013)	0.199 (0.013)
Female (X_2)	-4.347 (0.284)	-4.325 (0.283)
Midwest (X_3)	-1.587 (0.428)	
South (X_4)	-2.115 (0.405)	
West (X_5)	-0.948 (0.429)	
Adjusted R^2	0.080	0.075

(a) Calculate the R^2 for both models.

To calculate the R^2 knowing the Adjusted R^2, we can apply the formula:

$$R^2 = 1 - \left(\frac{N-k}{N-1}\right) \times (1 - \tilde{R}^2)$$

The two values of the R^2 for the two models will be respectively:

$$R^2_{(1)} = 1 - \left(\frac{5000-6}{5000-1}\right) \times (1 - 0.080) = 0.0806$$

$$R^2_{(2)} = 1 - \left(\frac{5000-3}{5000-1}\right) \times (1 - 0.075) = 0.0754$$

(b) Do men on average earn more or less than women? Is this difference significant?

$$\text{Difference} = -4.347$$

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0$$

The t-statistic is:

$$t = \frac{\beta_1}{SE(\beta_1)} = \frac{-4.347}{0.284} = -15.306$$

The difference is statistically significant (p-value close to zero).

(c) Calculate the 99% confidence interval for the difference between the hourly wages of a 25-year-old man and a 35-year-old man.

$$(1-\alpha)\%CI(\Delta AHE) = \Delta Age \times [\hat{\beta}_1 \pm z_{1-\alpha/2} \times SE(\hat{\beta}_1)]$$

$$99\%CI(\Delta AHE) = 10 \cdot [0.203 \pm z_{0.005} \cdot 0.013] = [1.6946; 2.3654]$$

(d) Perform a test to measure the significance of regional differences (significance level 1%).

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{vs} \quad H_1 : \exists \beta_j \neq 0$$

Knowing that:

$$F = \frac{R^2_{unrestricted} - R^2_{restricted}/q}{(1 - R^2_{unrestricted})/(n - k_{unrestricted})},$$

we can test the null hypothesis starting from the R^2 of the regressions.
For the non-restricted, the R^2 is that of column (1), while for the restricted we are going to consider the R^2 of column (2).
Considering the number of restrictions $q = 3$ and $n = 5,000$, it follows that

$$F = \frac{(0.0806 - 0.0754)/3}{(1 - 0.0806)/(5000 - 6)} = 9.4149$$

.

Given that the F-statistic is 9.4149, and the critical value is 3.78, we reject H_0. Regional differences are significant.
We can use also the Bonferroni test to test joint hypotheses on q coefficients starting from the t statistics relating to individual hypotheses but correcting the critical value as follows:

$$c = z_{1-\frac{\alpha/q}{2}}$$

H_0 is rejected if at least one of the individual t-ratios is, in absolute value, greater than the critical value c, adjusted as above.

In our case, H_0 is a joint hypothesis on 3 coefficients, therefore $q = 3$. Setting $\alpha = 1\%$, we have:

$$c = z_{1-\frac{\alpha/q}{2}} = z_{1-\frac{0.01/3}{2}} = z_{0.9983} = 2.93$$

The t-values are computed as the ratio of the estimated coefficient to its standard error (SE):

$$t = \frac{\beta_i}{SE(\beta_i)}$$

Thus, the t-values for each region are:
- Midwest:

$$t = \frac{-1.587}{0.428} \approx -3.71$$

- South:

$$t = \frac{-2.115}{0.405} \approx -5.22$$

- West:

$$t = \frac{-0.948}{0.429} \approx -2.21$$

The t-statistics of the single coefficients subject to the null hypothesis are $t_3 = -3.71$, $t_4 = -5.22$ and $t_5 = -2.21$. Since $|t_5| < c$ but $|t_3| > c$ and $|t_4| > c$, the null hypothesis is rejected.

2. Consider an autoregressive AR(1) model estimated on the Consumer Price Index (CPI) for the period from January 2000 to January 2009 (109 observations).

   We have the following data:

$$\sum y_t = 78.29, \quad \sum y_{t-1} = 81.96, \quad \sum y_{t-1}^2 = 226.838, \quad \sum y_t y_{t-1} = 140.378.$$

   (a) Calculate the estimates of the parameters β_0 and β_1.

The estimate of β_1 is given by:

$$\hat{\beta}_1 = \frac{E(y_t y_{t-1}) - E(y_t)E(y_{t-1})}{E(y_{t-1}^2) - [E(y_{t-1})]^2} = \frac{\frac{\sum y_t y_{t-1}}{T} - \bar{y}_t \bar{y}_{t-1}}{\frac{\sum y_{t-1}^2}{T} - \bar{y}_{t-1}^2}$$

Where for $T = 108$ (the 109 original observations excluding the first one for which we do not have the first lag):

$$\bar{y}_t = \frac{78.29}{108} = 0.725, \quad \bar{y}_{t-1} = \frac{81.96}{108} = 0.759$$

Calculation of $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\frac{140.378}{108} - 0.725 \cdot 0.759}{\frac{226.838}{108} - (0.759)^2} = \frac{1.524}{3.098} = 0.492$$

Estimate of the intercept β_0:

$$\hat{\beta}_0 = \bar{y}_t - \hat{\beta}_1 \bar{y}_{t-1} = 0.725 - 0.492 \cdot 0.759 = 0.352$$

Therefore, the estimated $AR(1)$ model is:

$$\hat{y}_t = 0.352 + 0.492 y_{t-1}$$

   (b) Knowing that this model has an $SER = 1.068$, calculate the BIC.

We know that:

$$\text{SER} = \sqrt{\frac{\sum \hat{u}_t^2}{n-k}}, \quad \text{where } \sum \hat{u}_t^2 = \text{RSS}$$

We calculate the sum of squared residuals (RSS):

$$\text{RSS} = \text{SER}^2 \times (n-k) = 1.068^2 \times (108 - 2) = 1.140624 \times 106 = 120.906144$$

$$\text{BIC} = \log\left(\frac{\text{RSS}}{T}\right) + \frac{k \log(T)}{T} = \log\left(\frac{120.91}{108}\right) + \frac{2\log(108)}{108} = 0.138$$

   (c) Calculate what the forecasts for the first three months of 2009 would be using an $AR(2)$ model estimated from:

$$\hat{y}_t = 0.537 + 0.604 y_{t-1} - 0.330 y_{t-2}$$

   and knowing that the observed values are:

   - November 2008: $y = -6.133$
   - December 2008: $y = -2.852$
   - January 2009: $y = 0.877$
   - February 2009: $y = 1.262$
   - March 2009: : $y = -0.342$

The $AR(2)$ model estimated up to December 2008 has parameters:

$$\hat{y}_t = 0.537 + 0.604 y_{t-1} - 0.330 y_{t-2}$$

The "one-step-ahead" forecasts for the first 3 months of 2009 will be:

$$\hat{y}_{\text{January 2009}} = 0.537 + 0.604 \cdot -2.852 - 0.330 \cdot -6.133 = 0.838$$

$$\hat{y}_{\text{February 2009}} = 0.537 + 0.604 \cdot 0.877 - 0.330 \cdot -2.852 = 2.008$$

$$\hat{y}_{\text{March 2009}} = 0.537 + 0.604 \cdot 1.262 - 0.330 \cdot 0.877 = 1.01$$

From this we will have the table of estimated values equal to:

(d) What will be the $RMSFE$ and with which method will it be possible to calculate it?

The Root Mean Squared Forecast Error ($RMS\hat{F}E_{OOS}$) is calculated as:

$$RMS\hat{F}E_{OOS} = \sqrt{\frac{1}{P}\sum_{t=1}^{P} \tilde{u}_t^2}$$

Where:

- $\tilde{u}_t = y_t - \hat{y}_t$ is the forecast error,
- P is the number of out-of-sample periods.

The table for the Out of Sample errors is:
Now, we compute the $RMS\hat{F}E_{OOS}$:

$$RMS\hat{F}E_{OOS} = \sqrt{\frac{1}{3}(0.001521 + 0.556516 + 1.828704)} = 0.892$$

3. Consider the following distributed lag (DL) model:

$$\hat{y}_t = \underset{(0.05)}{0.35} + \underset{(0.015)}{0.06}\, x_t + \underset{(0.02)}{0.08}\, x_{t-1} + \underset{(0.01)}{0.04}\, x_{t-2}$$

(a) Calculate the impact effect.

The impact effect is given by the coefficient of x_t, that is:

$$\hat{\beta}_1 = 0.06.$$

(b) Is the coefficient significant?

The t-statistic to verify significance is:

$$t = \frac{\hat{\beta}_1}{\mathrm{SE}(\hat{\beta}_1)} = \frac{0.06}{0.015} = 4.$$

The p-value is close to zero, so the impact effect is significant at the 1% significance level.

(c) Calculate the cumulative dynamic multiplier over two periods.

The cumulative dynamic multiplier over two periods is given by the sum of the coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$:

$$\hat{\delta}_3 = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = 0.06 + 0.08 + 0.04 = 0.18.$$

Therefore, the cumulative dynamic multiplier over two months is equal to 0.18.

(d) Are all dynamic multipliers significant?

We have already seen that the impact effect is significant. We need to calculate the significance for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$.

$$t = \frac{\hat{\beta}_0}{\mathrm{SE}(\hat{\beta}_0)} = \frac{0.35}{0.05} = 7.$$

$$t = \frac{\hat{\beta}_2}{\mathrm{SE}(\hat{\beta}_2)} = \frac{0.08}{0.02} = 4.$$

$$t = \frac{\hat{\beta}_3}{\mathrm{SE}(\hat{\beta}_3)} = \frac{0.04}{0.01} = 4.$$

The p-values for all coefficients are close to zero, so the intercept and all multipliers are significant.