ECOM20001: Econometrics 1
Final Exam 2019 S1 Solutions

Question 1: Multiple Choice (20 marks)

(1 mark per question)

1. Which of the following is not one the broad categories in which econometric analyses typically fall under?

   d. All of the above are broad categories for econometric analyses

2. Suppose you have a sample average of $\bar{Y} = 12$ with sample variance $s_Y^2 = 2$ and sample size $n = 25$. What is the 90% confidence interval for the population mean?

   b. [11.53,12.47]

3. You estimate a single linear regression:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

   and obtain a 99% confidence interval for β_1 of [-0.50,0.10]. Which of the following statements is necessarily true?

   d. None of the above are necessarily true

4. Which of the following is correct about the value of $\bar{R}^2$ in a multiple linear regression model?

   c. It can be negative

5. Suppose you run the following regression:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

   Further suppose there is an omitted variable Z_i that is positively correlated with Y_i and negatively correlated with X_i. Finally, suppose we obtained a regression coefficient $\hat{\beta}_1 < 0$. Based on the omitted variable Z_i, which of the following is true?

   b. $\hat{\beta}_1$ exhibits a negative omitted variable bias, causing it to be too large in magnitude

6. Suppose you run the following regression:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

   Which is the correct interpretation of β_1?

   a. A 1% increase in X_i has an associated 1% increase in Y_i

7. Suppose your OLS estimates suffer from imperfect multicollinearity, but where the OLS assumptions hold. Your OLS estimates will be:

   b. unbiased and inefficient

8. What problem does classical measurement error in a regressor create for regression analysis?

    b. it causes regression coefficients to be biased toward 0

9. Consider the following estimation results from a polynomial regression:

$$Y_i = \underset{(12.72)}{42} + \underset{(14.50)}{12.9}\, X_i - \underset{(17.92)}{0.34}\, X_i^2 + \underset{(8.22)}{0.01} X_i^3; \qquad \bar{R}^2 = 0.23, SER = 42.11$$

   What is the conclusion that can be drawn from these regression results?

    c. A scatter plot of Y_i on X_i is likely to exhibit less than 2 changes (or inflection points) in curvature in the relationship, which can give rise to imperfect multicollinearity in this regression

10. When testing a joint hypothesis with a multiple linear regression model, you should:

    b. use the F-statistic and reject at least one of the hypotheses if the statistic exceeds the critical value

Question 2: Short Answer Questions (20 marks)

a. Consider the following estimated single linear regression model using a sample of $n = 100$ observations:

$$Y = \underset{(0.2)}{10} - \underset{(0.1)}{0.5} \ln(X); \qquad \bar{R}^2 = 0.15$$

where the coefficient estimates' standard errors are in parantheses under the estimates. Interpret the regression coefficient on $\ln(X)$ and compute the 99% confidence interval on the coefficient. (6 marks)

The interpretation of the regression coefficient is a 1% increase in X is associated with a $0.5 \times 0.01 = 0.005$ unit decrease in Y

The 99% confidence interval, with a critical value of 2.58 for testing a two-sided hypothesis test that a regression coefficient equals 0. The 99% CI for the coefficient on X in the above regression is thus:

$$[-0.5 - 2.58 * 0.1, -0.5 + 2.58 * 0.1] = [-0.758, -0.242]$$

b. Write down the formula of the Homoskedasticity-only F-statistic, explain the steps you would take to compute it, and state the null and alternative hypotheses for the test that corresponds to the statistic. Briefly comment on the connection between the regression R^2 and F-statistic for joint hypothesis testing highlighted by the Homoskedasticity-only F-statistic. (6 marks)

The Homoskedasticity-only F-statistic is:

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

where n is the number of observations, k is the number of regressors, q is the number of restrictions. To compute the two SSR's in the formula we have to run two different regressions:

- $SSR_{unrestricted}$ is the unrestricted regression SSR (with k regressors) not imposing any restrictions,
- $SSR_{restricted}$ is the regression SSR (with $k - q$ unrestricted regressors) that imposes the q restrictions under the null hypothesis H_0

Specifically the null hypothesis that is tested with this F-statistic is:

$$H_0 : \beta_j = \beta_{j,0},\ \beta_m = \beta_{m,0}, \ldots \text{ for a total of } q \text{ restrictions}$$

with alternative:

$$H_1 : \beta_j \neq 0 \text{ for at least one } j, j = 1, \ldots, k$$

Intuitively, if there is a large increase in SSR (or reduction in the R^2) when the q restrictions are imposed, the more likely the null hypothesis is to be false (and rejected using the F-statistic).

c. Suppose you estimate a single linear regression model $Y = \beta_0 + \beta_1 X + u$ and obtain an OLS regression coefficient estimate of $\hat{\beta}_1$. Holding all other aspects of the data fixed except sample n size, explain why you are more likely to find that the OLS regression coefficient estimate $\hat{\beta}_1$ is statistically significantly different from 0 using a two-sided hypothesis test as n grows. Make explicit use of appropriate equations and/or formulas in answering your question. (8 marks)

The variance error formula for the OLS slope estimate is :

$$\widehat{\sigma^2_{\hat{\beta}_1}} = \frac{1}{n}\frac{var((X_i - \bar{X})\hat{u}_i))}{(var(X_i))^2}$$

with the standard error being:

$$SE(\hat{\beta}_1) = \sqrt{\widehat{\sigma^2_{\hat{\beta}_1}}}$$

and the t-statistic formula for the test of the null that β_1 equals 0 is given by:

$$t^{act} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$$

Now suppose we hold the OLS estimate fixed at $\hat{\beta}_1$, and consider only an increase in sample size n. This causes $SE(\hat{\beta}_1) = \sqrt{\widehat{\sigma^2_{\hat{\beta}_1}}}$ to fall as n is in the denominator of $\widehat{\sigma^2_{\hat{\beta}_1}}$. This causes t^{act} to increase as $SE(\hat{\beta}_1)$ is in the denominator of t^{act}. Now, given that the p-value for the hypothesis test stated in the question is computed as $2\Phi(-|t^{act}|)$, this implies that increasing n will cause $-|t^{act}|$ to become smaller, and hence the p-value to become smaller. As the p-value becomes smaller, by definition, you are more likely to find $\hat{\beta}_1$ is statistically significantly different from 0.

Question 3: MetricsBars (10 marks)

A chocolate bar retailer asks you to evaluate the impact of a marketing campaign they ran for their MetricsBars chocolate bar. They provide you with a raw dataset called **marketing.csv**, which contains the following variables:

> $sales_i$: sales of MetricsBars in market i in dollars (\$)
>
> $campaign_i$: a dummy variable equalling 1 if the marketing campaign was run in market i, and is equal to 0 otherwise.
>
> $income_i$: average individual income in market i in dollars (\$)

These data are provided for a sample of $n = 1000$ markets, where 500 markets were randomly chosen to have the marketing campaign, and the other 500 markets did not have the marketing campaign. The sample average of $sales_i$ is \$1,000,000 and the sample average of $income_i$ is \$50,000.

MetricBars wants to inform the following question:

> *If household income increases from \$50,000 to \$70,000, how much more effective will the marketing campaign be at increasing MetricsBars sales in percentage terms?*

Starting from the raw data in **marketing.csv**, write down the pseudo-code in R you would develop to provide a relevant 95% confidence interval that empirically informs this question. List the steps you would include in your code, and if it helps in describing your answer, you may state explicit R code though this is not necessary for obtaining full marks. Be precise and explicitly describe any variable scaling you would use, regressions run in your code, hypothesis tests required, test statistics used, or any other calculations necessary for computing the 95% confidence interval.

Steps in the pseudo-code are as follows:

1. Load the AER() package, load the data, and rescale average houshold income by \$10000: $income_i = income_i/10000$

2. Compute a new logarithmic variable:

$$log\_sales_i = \ln(sales_i)$$

3. Compute a new interaction variable:

$$campaign\_income_i = campaign_i \times income_i$$

4. Run the following regression, allowing for heteroskedastic standard errors:

$$log\_sales_i = \beta_0 + \beta_1 campaign_i + \beta_2 campaign\_income_i + \beta_3 income_i + u_i$$

5. From the estimated log-linear regression, the percentage change in sales associated with the marketing campaign when household earnings goes from \$50,000 to \$70,000 is computed in 2 steps. Impact of marketing campaign at \$50,000 is:

$$\Delta\hat{Y}_1 = (\beta_0 + \beta_1 1 + \beta_2 5 + \beta_3 5) - (\beta_0 + \beta_3 5) = \beta_1 + \beta_2 5$$

and impact of marketing campaign at \$70,000 is:

$$\Delta\hat{Y}_2 = (\beta_0 + \beta_1 1 + \beta_2 7 + \beta_3 7) - (\beta_0 + \beta_3 7) = \beta_1 + \beta_2 7$$

where note the 5 and 7 in the calculation corresponds to the \$10,000 scaling of $income_i$ above. So the change in the effectiveness of the marketing campaign when income goes from \$50,000 to \$70,000 is:

$$\Delta\hat{Y}_2 - \Delta\hat{Y}_1 = 2\beta_2$$

6. To compute the standard error of $\Delta\hat{Y}$, we first run the following joint-hypothesis test:

$$H_0 : 2\beta_2 = 0; \qquad \text{vs. } H_1 : 2\beta_2 \neq 0$$

We compute the F-statistic from R associated with this test, call it F. With this, we can compute $SE(\Delta\hat{Y})$ as:

$$SE(\Delta\hat{Y}) = \frac{|\Delta\hat{Y}|}{\sqrt{F}}$$

7. Finally, we can compute the 95% CI of the change in the impact of the marketing campaign when income goes from \$50,000 to \$70,000 as:

$$[\Delta\hat{Y} - 1.96 \times SE(\Delta\hat{Y}), \Delta\hat{Y} + 1.96 \times SE(\Delta\hat{Y})]$$

Question 4: Wages, Experience, and Education (20 Marks)

The Department of Jobs and Small Business has approached you to study the relationship between wages, experience, and education. They have provided you a dataset which contains the following information from a random sample of $n = 801$ individuals:

$wage_i$: hourly wage earned by individual i in dollars ($)

$exper_i$: experience of individual i working measured as the number of years they have been working in the labour market

$degree_i$: a dummy variable equalling 1 if individual i has a university degree and equals 0 if they do not have a university degree.

age_i: age of individual i in years

$urban_i$: dummy variable equalling 1 if individual i lives in an urban location, and equals 0 if they live in a regional location.

The Department wants to understand how wages change with work experience in the labour market. Figure 1 on the next page produces summary statistics for these data, and regression output from R for three different regressions that focus on the relationship between wages and experience. Based on this output, answer the following questions. Throughout assume a 5% level of significance in conducting hypothesis tests.

a. What percentage of individuals in the data set have a university degree? (2 marks)

   71.4% of individuals in the data set have a university degree.

b. Interpret the statistical significance, sign and magnitude of the regression coefficient estimate on $exper_i$ in Regression 1. (3 marks)

   A 1-year increase in work experience is associated with a $0.41 increase in an individual's hourly wage. This relationship is statistically significant at the 5% level as it has a p-value of $0.007 < 0.05$ (could also compare the 2.7277 t-statistic to the 1.96 critical value).

c. The regression coefficient on $exper_i$ changes substantially between Regressions 1 and 2. Carefully explain what might drive this large change in the regression coefficient on $exper_i$ between Regressions 1 and 2. (4 marks)

   There are two important correlations to consider when assessing the omitted variable bias:

   - age_i is likely positively correlated with $exper_i$ as older people tend to have more work experience, call this correlation $+$
   - age_i is likely positively correlated with $wage_i$ as older people tend to have higher earnings over time, call this correlation $+$

   Therefore, the sign of the bias when age_i is in the error term in Regression 1 is $sign(bias) = + \times + = +$. That is, by not controlling for age_i we have positive upward bias in the coefficient on $exper_i$ in Regression 1. Once we remove this bias in Regression 2 by controlling for age_i, we see as argued that the coefficient on $exper_i$ becomes smaller.

d. Interpret the statistical significance, sign and magnitude of the regression coefficient estimate on $degree_i$ in Regression 3. (3 marks)

Holding experience, age, and urban location fixed, relative to an individual without a university degree, and individual with a university degree is expected to have a \$18.05 higher hourly wage. The difference is statistically significantly different from 0 at the 5% level based on the p-value which is less than 0.01 (2 e-16).

e. Based on the regression output in Figure 1, is there evidence of a nonlinear relationship between $wage_i$ and $exper_i$? Carefully state the null and alternative for the relevant hypothesis test for conducting this test, and highlight what regression coefficient estimate(s), t-statistic(s), and p-value(s) in Figure 1 allow you to conduct such a test for a nonlinear relationship. (3 marks)

Holding other factors fixed, the coefficient on $exper\_sq_i$ in the quadratic regression in Regression 3 is what determines if there is a nonlinear relationship between $wage_i$ and $exper_i$. The relevant hypothesis to test is the null that this coefficient equals 0 against the two-sided alternative that it is not equal to 0. From the regression output, we see that this test has a corresponding p-value of $0.021 < 0.05$ which implies it is statistically significant at the 5% level. Therefore, statistically, the regression implies a nonlinear relationship between $wage_i$ and $exper_i$.

f. Based on Regression 3, holding age and urban status fixed, for an individual starting with 0 years experience and without a university degree, how many years will they have to work in the labour market to "catch up" in terms of expected wages to an individual with 0 years experience but with a university degree? (5 marks)

Based on the results in Regression 3, wages grow with experience at a rate of 0.38 $exper$ - 0.014 $exper^2$. Moreover, holding other factors fixed, an individual with a university degree is expected to have a \$18.05 higher wage than an individual without a university degree. Therefore, starting from no experience, the number of years experience required for someone without a degree to catch an individual with a degree and no experience solves the following equation:

$$18.05 = 0.38 exper - 0.014 exper^2$$

In this particular case, however, the regression yields a quadratic formula that does not have an real solutions. This implies that the model predicts that starting from no experience, an individual without a degree will never catch and individual with a degree and no experience.

Figure 1: Summary Statistics and Estimation Results for the Wage Regressions

```
##### SUMMARY STATISTICS #####

> mydata=read.csv(file="q4_wage_dat.csv")

> summary(mydata)
      wage           exper            age            degree           urban       
 Min.   :34.85   Min.   : 8.00   Min.   :22.75   Min.   :0.0000   Min.   :0.0000  
 1st Qu.:41.62   1st Qu.:13.00   1st Qu.:28.50   1st Qu.:0.0000   1st Qu.:0.0000  
 Median :56.75   Median :14.00   Median :30.25   Median :1.0000   Median :0.0000  
 Mean   :52.54   Mean   :14.39   Mean   :30.31   Mean   :0.7141   Mean   :0.4931  
 3rd Qu.:58.49   3rd Qu.:16.00   3rd Qu.:32.00   3rd Qu.:1.0000   3rd Qu.:1.0000  
 Max.   :63.86   Max.   :22.00   Max.   :37.75   Max.   :1.0000   Max.   :1.0000  


##### REGRESSION 1 #####

> reg1=lm(wage~exper,data=mydata)

> coeftest(reg1, vcov = vcovHC(reg1, "HC1"))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 46.61942    2.20113 21.1798 < 2.2e-16 ***
exper        0.41174    0.15095  2.7277  0.006517 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



##### REGRESSION 2 #####

> reg2=lm(wage~exper+age,data=mydata)

> coeftest(reg2, vcov = vcovHC(reg2, "HC1"))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)    
(Intercept) 37.80714    3.68366 10.2635 < 2.2e-16 ***
exper        0.10760    0.17977  0.5985  0.549657    
age          0.43513    0.14670  2.9660  0.003107 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



##### REGRESSION 3 #####

> mydata$exper_sq=mydata$exper*mydata$exper

> reg3=lm(wage~exper+exper_sq+age+degree+urban,data=mydata)

> coeftest(reg3, vcov = vcovHC(reg3, "HC1"))

t test of coefficients:

              Estimate Std. Error  t value Pr(>|t|)    
(Intercept) 21.0229602  1.3044264  16.1166  < 2e-16 ***
exper        0.3823205  0.1760285   2.1719  0.03016 *  
exper_sq    -0.0143891  0.0062149  -2.3153  0.02085 *  
age          0.5014108  0.0170532  29.4028  < 2e-16 ***
degree      18.0537071  0.0786755 229.4705  < 2e-16 ***
urban        1.9569213  0.0727571  26.8966  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Question 5: Understanding the Demand for Sunscreen (20 Marks)

The Department of Health wants to understand the relationship between sunscreen demand month-to-month, and the number of tourists visiting Australia. To conduct this analysis, you have been provided the following dataset:

$sunscreen_t$: total bottles of sunscreen sold in terms of 1000s of bottles in month t

$lag\_sunscreen_t$: one-month lag of $sunscreen_t$ (e.g., $lag\_sunscreen_t = sunscreen_{t-1}$)

$tourists_t$: total number of tourists visiting Australia in terms of 1000s of tourists in month t

$lag\_tourists_t$: one-month lag of $tourists_t$ (e.g., $lag\_tourists_t = tourists_{t-1}$)

$month_t$: month of year for month t, taking on one of 12 values in the following list: {Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec}

You have this information for $T = 135$ months. Figures 2 and 3 on the next two pages respectively present time series plots and regression output generated by R that analyzes these data. Based on this information, please answer the following questions:

a. Based on Figure 2 explain whether $sunscreen_t$ appears to be stationary and whether it exhibits seasonality. Similarly, based on Figure 2 explain whether $tourists_t$ appears to be stationary and whether it exhibits seasonality. (2 marks)

   Both time series plots do not exhibit a persistent trend, which suggests both are stationary. Also both time series plots exhibit cyclical patterns, which points to strong seasonality. The seasonality is somewhat clearer for $tourists_t$, but we also see regular spikes in sunscreen sales.

b. Explain what the residuals from in Regression 1 in Figure 3 would represent compared to the raw $sunscreen_t$ time series in the data set. (2 marks)

   Note: For the remainder of the question, note that Figure 3 is displayed over pages 11 and 12 below.

   Just before Regression 1 in the R code, a set of month-of-year time dummies is being created. Therefore, the residuals from this regression would remove any month-of-year seasonal patterns in $sunscreen_t$, implying that this time series would exhibit far less cyclicality than the raw time series for $sunscreen_t$.

c. Explain how Regression 1 avoids a potential dummy variable trap based on the dummy variables created in the code, and the dummy variables included in the regression. (2 marks)

   The regression drops the December dummy variable to avoid the dummy variable trap via the constant in the regression. If all 12 month dummies were included, they would be colinear with the constant.

d. Suppose that the last data point at $T = 135$ in the sample is January, has a value of $sunscreen_{135} = 8$, and a value of $tourists_{135} = 12$. Based on Regression 2 in Figure

3, what would be your out-of-sample forecast for $sunscreen_t$ in period $T = 136$ and the 95% forecast interval assuming IID normal errors in the regression equation in Regression 2? Work with 3 digits after the decimal in conducting your calculations. (4 marks)

From the regression equation, the out-of-sample forecast would be:

$$\widehat{sunscreen}_{T+1} = 0.941 + 0.178 \times 8 + 0.118 \times 12 + 0.028 = 3.908$$

The SER in the regression is 0.096, which implies a 95% forcast interval of:

$$[3.908 - 1.96 * 0.096, 3.908 + 1.96 * 0.096] = [3.719, 4.096]$$

e. Compute the Bayes-Schwartz Information Criterion for Regressions 1 and 2 in Figure 3 and explain which is the preferable time series model based on this criterion. Work with 3 digits after the decimal in conducting your calculations. (6 marks)

The BIC for the two regressions, with $K = 12$ and $K = 14$ parameters respectively, can be computed as:

- Regression 1:
$$BIC(12) = \ln\left[\frac{SSR(12)}{135}\right] + 12\frac{\ln(135)}{135}$$
- Regression 2:
$$BIC(14) = \ln\left[\frac{SSR(14)}{135}\right] + 14\frac{\ln(135)}{135}$$

We can compute the SSRs from the SER reported in the regression using the SER formula:

$$SER = s_{\hat{u}} = \sqrt{s_{\hat{u}}^2}; \qquad s_{\hat{u}}^2 = \frac{SSR}{n-k-1}$$

implying

$$SSR = SER^2 \times (n-k-1)$$

and thus from the regression output for Regressions 1 and 2

$$SSR(12) = 0.1 \times 0.1 \times (135 - 11 - 1) = 1.23$$

and

$$SSR(14) = 0.096 \times 0.096 \times (135 - 13 - 1) = 1.115$$

So the BICs are computed as:

- Regression 1:
$$BIC(12) = \ln\left[\frac{1.23}{135}\right] + 12\frac{\ln(135)}{135} = -4.26$$
- Regression 2:
$$BIC(14) = \ln\left[\frac{1.115}{135}\right] + 14\frac{\ln(135)}{135} = -4.29$$

which implies Regression 2 is the preferred model according to the BIC since it yields a smaller value.

f. Conduct a Granger Causality test to determine whether $tourists_t$ Granger Causes $sunscreen_t$ based on the regression results in Regression 2 in Figure 3. Assume a 5% level of significance in conducting the test. Work with 3 digits after the decimal in conducting your calculations. (4 marks)

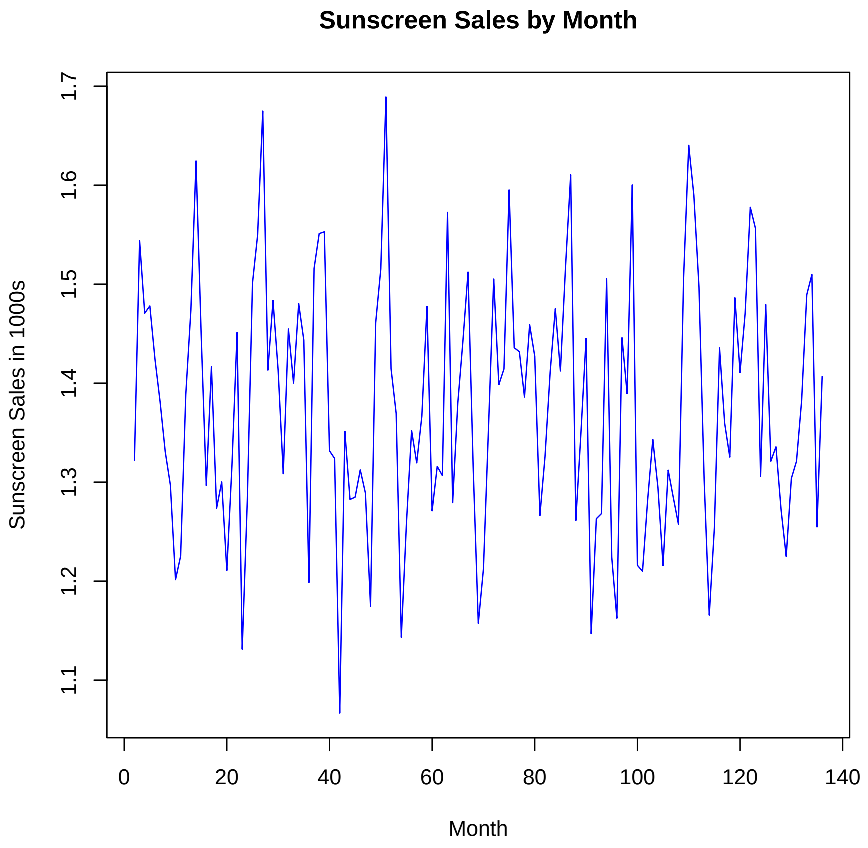
Letting β_2 be the coefficient on $lag\_tourists_t$, the Granger Causality test in this application has the following null and alternative hypothesis:

$$H_0 : \beta_2 = 0 \text{ vs. } H_1 : \beta_2 \neq 0$$

From the regression results, we see this has a t-statistic of 2.4867, implying a F-statistic for the test (which is the square of the t-statistic) of $F = 2.487 \times 2.487 = 6.185$. The degrees of freedom for the test is $q = 1$ and $(T - \max\{p, q_1, \ldots, q_k\} - p - \sum_{\ell=1}^{k} q_\ell - 1 = 135 - 1 - 13 - 1 = 120)$.

From the critical values table for the 95th percentile of the F(1,120) distribution at the end of the exam paper, we have a critical value of 3.92 for the test. Since $6.185 > 3.92$, we conclude from the test at the 5% level that $tourists_t$ does in fact "Granger cause" $sunscreen_t$.

Figure 2: Time Series Plots of Sunscreen Sales and Tourists by Month



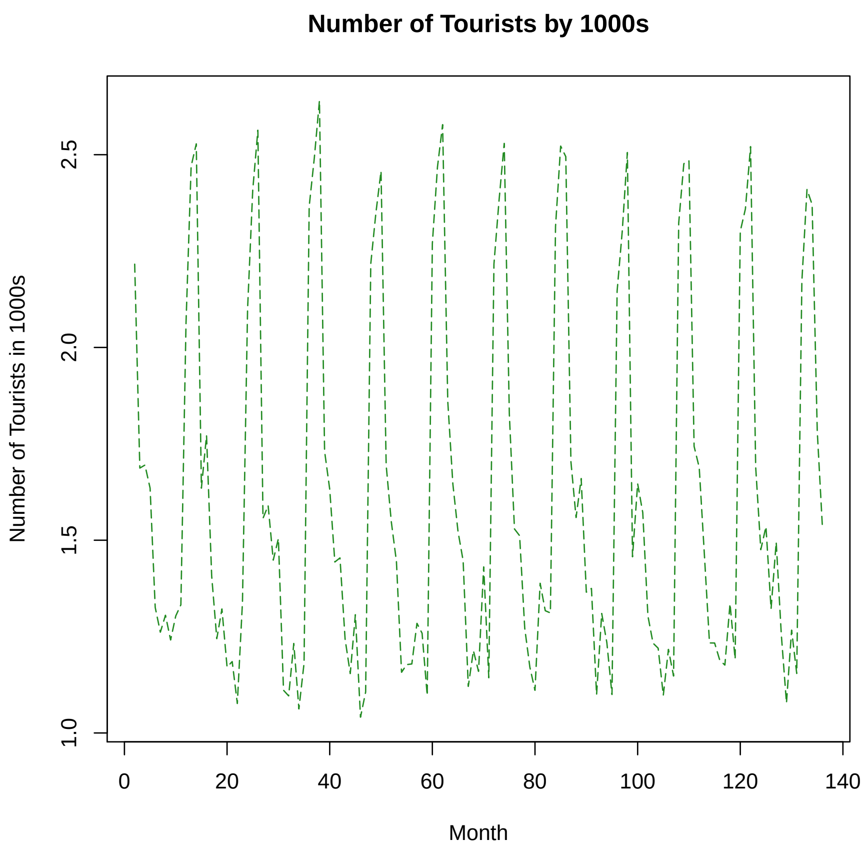


Figure 3: Estimation Results for the Time Series Regressions

```
> mydata=read.csv(file="q5_sunscreen_dat.csv")

##### CREATING DUMMY VARIABLES #####

> mydata$dJan=as.numeric(mydata$month=="Jan")
> mydata$dFeb=as.numeric(mydata$month=="Feb")
> mydata$dMar=as.numeric(mydata$month=="Mar")
> mydata$dApr=as.numeric(mydata$month=="Apr")
> mydata$dMay=as.numeric(mydata$month=="May")
> mydata$dJun=as.numeric(mydata$month=="Jun")
> mydata$dJul=as.numeric(mydata$month=="Jul")
> mydata$dAug=as.numeric(mydata$month=="Aug")
> mydata$dSep=as.numeric(mydata$month=="Sep")
> mydata$dOct=as.numeric(mydata$month=="Oct")
> mydata$dNov=as.numeric(mydata$month=="Nov")
> mydata$dDec=as.numeric(mydata$month=="Dec")



##### REGRESSION 1 ESTIMATES AND MODEL FIT #####

> reg1=lm(sunscreen~dJan+dFeb+dMar+dApr+dMay+dJun+dJul+dAug+dSep+dOct+dNov,data=mydata)

> coeftest(reg1, vcov = vcovHC(reg1, "HC1"))

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  1.318972   0.036075 36.5615 < 2.2e-16 ***
dJan         0.134728   0.040205  3.3510 0.0010698 **
dFeb         0.174234   0.048269  3.6096 0.0004446 ***
dMar         0.238514   0.048816  4.8860 3.137e-06 ***
dApr         0.041750   0.044544  0.9373 0.3504455
dMay         0.065165   0.044312  1.4706 0.1439548
dJun        -0.013259   0.053756 -0.2467 0.8055894
dJul         0.012884   0.046883  0.2748 0.7839271
dAug         0.011853   0.043214  0.2743 0.7843192
dSep        -0.036224   0.041502 -0.8728 0.3844647
dOct         0.026185   0.046862  0.5588 0.5773346
dNov         0.012821   0.049966  0.2566 0.7979189
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


# Regression 1 Fit
> summary(reg1)
# (note: regression output for reg1 with
# homoskedastic standard errors omitted)

Residual standard error: 0.1 on 123 degrees of freedom
Multiple R-squared:  0.4201,      Adjusted R-squared:  0.3682
F-statistic:   8.1 on 11 and 123 DF,  p-value: 1.63e-10
```

Estimation Results for the Time Series Regressions (Figure 3 continued)

```
##### REGRESSION 2 ESTIMATES AND MODEL FIT #####

> reg2=lm(sunscreen~lag_sunscreen+lag_tourists+
+           dJan+dFeb+dMar+dApr+dMay+dJun+dJul+dAug+dSep+dOct+dNov,data=mydata)

> coeftest(reg2, vcov = vcovHC(reg2, "HC1"))

t test of coefficients:

                Estimate Std. Error t value  Pr(>|t|)
(Intercept)    0.9413073  0.1013951  9.2836 8.207e-16 ***
lag_sunscreen  0.1776939  0.0767075  2.3165   0.02221 *
lag_tourists   0.1183732  0.0476024  2.4867   0.01426 *
dJan           0.0145053  0.0609275  0.2381   0.81223
dFeb           0.0281326  0.0610873  0.4605   0.64596
dMar           0.0560163  0.0787568  0.7113   0.47829
dApr          -0.0582770  0.0487922 -1.1944   0.23466
dMay           0.0104385  0.0456310  0.2288   0.81944
dJun          -0.0606499  0.0517203 -1.1727   0.24324
dJul           0.0011059  0.0449387  0.0246   0.98041
dAug           0.0049781  0.0426757  0.1166   0.90733
dSep          -0.0348500  0.0392062 -0.8889   0.37582
dOct           0.0310750  0.0436303  0.7122   0.47769
dNov           0.0057216  0.0480600  0.1191   0.90543
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


# Regression 2 Fit
> summary(reg2)
# (note: regression output for reg2 with
# homoskedastic standard errors omitted)

Residual standard error: 0.09623 on 121 degrees of freedom
Multiple R-squared:  0.4721,    Adjusted R-squared:  0.4153
F-statistic: 8.323 on 13 and 121 DF,  p-value: 8.566e-12
```