ECONOMETRICS LABS

to accompany *Introduction to Econometrics* by Gary Koop

Prof. Ross McKitrick
Department of Economics and Finance
University of Guelph
ross.mckitrick@uoguelph.ca

UNIT 1: OVERVIEW OF ECONOMETRICS

Readings: Chapter 1

Topics
Types of data
Transformations of data
Graphs: time series, histograms, XY scatters
Estimation—choosing best value of a parameter
Least squares principle and the mean
Descriptive stats: mean, MSE, Variance, Std Dev
Sample versus population
Expected value
Population variance
Correlation
Population covariance

Questions

1. Ensure you have worked through Sections 1 to 3 in the R Studio Tutorial.

2. (Expected value) Suppose the forest service in a certain region tabulates data on July forest fires and obtains the following results.

Number	Probability
0	0.16
1	0.24
2	0.23
3	0.18
4	0.09
5	0.06
6	0.04
7 or more	0.00

   If we denote the number of fires as X, what is $E(X)$, the expected number of forest fires in that region in July?

3. (Variance) Compute $E(X-\mu)^2$ where μ is the expected number of fires $E(X)$ as computed in Question 1.

4. (Variance) Compute $E(X^2)-\mu^2$ and compare to your answer in Question 2. Hint: the answer should be 2.5804 each time.

5. Read Appendix B Definition B.8 and Theorem B.2 (pp. 341—342) in the textbook.

6. (Population and sample statistics) Review and memorize the following concepts. There is a blank table on the next page for you to practise with.

Population $X_1, \dots, X_N$	**Sample** $x_1, \dots, x_n$
The mean μ = the expected value $E(X)$	The sample mean is $\bar{x} = \frac{1}{n}\Sigma x_i$
The variance $\sigma^2 = E(X-\mu)^2$	The sample variance is $s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2$
The variance of the mean is $E(\bar{X}-\mu)^2 = \frac{\sigma^2}{N}$	The variance of the sample mean is $\frac{s^2}{n}$.
The standard deviation of X is $\sqrt{\sigma^2} = \sigma$.	The sample standard deviation is $\sqrt{s^2} = s$
The standard error of the mean is $\frac{\sigma}{\sqrt{N}}$	The standard error of the sample mean is $\frac{s}{\sqrt{n}}$
The covariance between two variables X and Y is $\sigma_{XY} = E(X-\mu_X)(Y-\mu_Y)$.	The sample covariance between two variables X and Y is $s_{xy} = \left(\frac{1}{n-1}\right)\Sigma_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$.

Population $X_1, \ldots, X_N$	**Sample** $x_1, \ldots, x_n$
The mean μ = the expected value	The sample mean is $\bar{x} =$
The variance $\sigma^2 =$	The sample variance is $s^2 =$
The variance of the mean is $E(\bar{X} - \mu)^2 =$	The variance of the sample mean is
The standard deviation of X is	The sample standard deviation is $\sqrt{s^2} =$
The standard error of the mean is	The standard error of the sample mean is
The covariance between two variables X and Y is $\sigma_{XY} =$	The sample covariance between two variables X and Y is $s_{xy} =$

7. Review the following formulas. If X and Y are both random variables and a and b are constants:

$$E(X + Y) = E(X) + E(Y)$$

$$E(a + bY) = a + bE(Y)$$

$$E(aX + bY) = aE(X) + bE(Y)$$

$$cov(X, Y) = \sigma_{xy} = E(X - \mu_x)(Y - \mu_y)$$

$$corr(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

$$var(a + X) = var(X)$$

$$var(bX) = b^2 var(X)$$

$$var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2(ab)cov(X, Y)$$

8. Suppose X and Y are independent and each has a variance of 20. What is $var(X + Y)$?

9. Suppose $Y = X + 2$ and $var(X) = 7$. What is $var(Y)$?

10. Suppose $Y = 2X$ and $var(X) = 1$. What is $var(Y)$?

11. Suppose $var(X) = 4$, $var(Y) = 3$ and $cov(X, Y) = 3.6$. What is $var(9X + 2Y)$?

UNIT 2: INTRO TO SIMPLE REGRESSION

Readings: Chapters 2.1, 2.2

Topics
Models
Regression
Estimation and optimal fit: least squares rule
Interpreting parameters
Measuring the fit: SSR, TSS, RSS
R^2
Confidence intervals
Hypothesis tests
t and F statistics; *p* value

Questions

1. Suppose X_i= observations on house prices and $Y_i = 7$ (i.e. a constant) for a sample $i = 1, \dots, N$. Prove that the covariance of *X* and *Y* is zero.

2. Why do we call the procedure for fitting a line through observations the "least squares" rule?

3. Prove that the least squares estimator of the mean is $\bar{X} = \frac{1}{N}\Sigma_{i=1}^{N} X_i$. To do the proof, start with the fact that the mean $\bar{X}$ is meant to approximate or represent each observation X_i in the data set, but there is a residual $\epsilon_i = X_i - \bar{X}$ representing the difference between each variable and the mean. Write out the formula for the sum of the squared residuals and use calculus to find the expression for $\bar{X}$ that minimizes it.

4. Suppose we have a variable Y_i measuring unemployment rates across *N* cities, and we regress it on a constant:

$$Y_i = \alpha + \epsilon_i.$$

   What will R^2 be?

5. Suppose we have two variables: X_i = population in country *i*, and $Y_i = 2X_i$. If the variance of X_i = 9.1, what is the variance of Y_i? What is the correlation between *X* and *Y*?

6. Use the dataset orange.csv (from the textbook data library) and the following R program to construct the variables *organic*, *regular* and *new*.

```
rm(list=ls(all=TRUE))
fruit = read.table("orange.csv", sep=",", skip=2)
organic = fruit$V1
regular = fruit$V2
new = log(organic) + sqrt(regular)
```

Run a regression of *regular* on *new* using the `lm()` command. What is the slope coefficient? Verify the t statistic by computing it yourself using the coefficient estimate and the standard error.

7. Suppose we have a regression model $Y_i = \alpha + \beta X_i + \epsilon_i$. Write out the expression for the fitted regression line.

8. If ϵ_i is the *error* term, what do we call $\hat{\epsilon}_i$?

9. For the regression model in question 7, the derivative $\frac{d\hat{Y}_i}{dX_i} = \hat{\beta}$ implies (choose the correct answer)

   a) the slope coefficient measures the marginal effect of X on Y
   b) the slope coefficient measures the goodness of fit of the regression
   c) the dependent variable Y is independent of the error term
   d) the explanatory variable X is unrelated to the OLS estimate.

10. Suppose we have a regression model $W_i = \alpha + \beta Z_i + \epsilon_i$ and we estimate it using a sample of N observations. Write out the formulas for
    a) The Total Sum of Squares (TSS)
    b) The Regression Sum of Squares (RSS)
    c) The Sum of Squared Residuals (SSR)
    d) R^2
    e) The F statistic

UNIT 3: INTRO TO MULTIPLE REGRESSION

Readings: Chapters 2.3, 2.4

Topics
The linear model
Omitted variable bias
Multicollinearity
Dummy variables

Questions

1. Do Textbook Chapter 2 Question 1 (omit 1c).

2. Suppose we have a dependent variable Y_i whose mean is zero. Then suppose we regress Y_i on X_i and find the slope coefficient is zero. Show that R^2 must be zero.

3. A data set consists of observations on education spending per capita (Ed_i), Gross Domestic Income per capita ($inc\_per\_cap_i$) and total population (POP_i) for N countries. We run the following regression:

$$Ed_i = \alpha + \beta_1 \times inc\_per\_cap_i + \beta_2 \times POP_i + \epsilon_i.$$

The following results are obtained using R:

```
Call:
lm(formula = ed ~ inc_per_cap + POP)

Residuals:
    Min      1Q  Median      3Q     Max
-701.47  -26.68   13.07   47.38  551.11

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -79.447252  67.865310  -1.171    0.250
inc_per_cap   0.058773   0.003991  14.726   <2e-16 ***
POP           0.458092   1.830679   0.250    0.804
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 213 on 35 degrees of freedom
Multiple R-squared:  0.8772,    Adjusted R-squared:  0.8702
F-statistic:   125 on 2 and 35 DF,  p-value: < 2.2e-16
```

(a) What is the sample size? What is the value of the constant term?
(b) What is the marginal effect of a one unit increase in population on Ed_i?
(c) Is the effect significantly different from zero?

4. In the above regression, what is the F statistic and what does it imply?

UNIT 4: ASSUMPTIONS BEHIND SIMPLE REGRESSION

Readings: Chapters 3.1—3.3

Topics
Probability distributions and density functions
Standard normal and related distributions: χ^2, t and F
Classical Assumptions

Questions

1. If $E(Y) = \mu$ and $var(Y) = \sigma^2$ and $Z = \frac{Y-\mu}{\sigma}$, prove that $E(Z) = 0$ and $var(Z) = 1$.
2. If Z follows a standard normal distribution, evaluate the following probabilities:
   (a) Pr $(Z \geq 0)$
   (b) Pr $(Z \geq 1)$
   (c) Pr $(Z < 1)$
   (d) Pr$(Z \leq 1.5)$
   (e) Pr $(-0.5 \leq Z \leq 0.5)$
   (f) Pr $(-0.6 \leq Z \leq 0.8)$

3. Explain in words the meaning of the Classical Assumption $cov(\epsilon_i, \epsilon_j) = 0\ for\ i \neq j$.

4. "We expect Y_i to lie on the regression line." Which Classical Assumption expresses this?

5. Does the fact that an observation Y_i does not lie on the regression line mean that assumption was violated? Explain.

6. Why is it important to spell out all the assumptions about the error term ϵ_i?
   a) Because they determine how the observations X_i are distributed.
   b) Because they determine how the estimated coefficient $\hat{\beta}$ is distributed.
   c) Because they determine how the observations Y_i are distributed.
   d) Because they determine whether the regression model is linear.

7. What parameters are needed in order to fully characterize the Normal distribution?

8. Suppose a random variable X_i follows a Normal distribution. What is the formula for its estimated variance?

9. Recall that a function of random variables is also a random variable. The following results are not in the Koop textbook but are presented to help you understand where the other distributions come from that we use in econometrics.

   We will make use of three distributions referred to as χ^2 ("chi-squared"), t and F. Unlike the standard normal distributions, these are all families of distributions that vary depending on the sample size. In order to use the correct version of each distribution we have to specify the sample size N and the number of estimated parameters k. The difference N-k is called the *degrees of freedom*.

Suppose Z follows a standard normal distribution. Then Z^2 is also a random variable and it follows a distribution called χ^2 with 1 degree of freedom.

Suppose we have *n* independent standard normal variables denoted Z_i. If $X = \Sigma_{i=1}^{n} Z_i^2$ (i.e. the sum of the squared Z's) then X is also a random variable and it follows the χ^2 distribution with *n* degrees of freedom.

Suppose that Z is a standard normal variable and X is χ^2 with *n* degrees of freedom and the two are independent. Then the ratio $t = \frac{Z}{\sqrt{X/n}}$ is a random variable and it follows a *t* distribution with *n* degrees of freedom.

Suppose that X_1 follows a χ^2 distribution with k_1 degrees of freedom and X_2 follows a χ^2 distribution with k_2 degrees of freedom. Then the ratio

$$F = \left(\frac{X_1}{k_1}\right) / \left(\frac{X_2}{k_2}\right)$$

is also a random variable and it follows an *F* distribution with k_1 and k_2 degrees of freedom.

Given all the above, is t^2 a random variable, and if so, what distribution does it follow?

UNIT 5: PROPERTIES OF OLS ESTIMATORS & THE GAUSS-MARKOV THEOREM

Readings: Chapters 3.4, App 3-1

Topics
Unbiasedness
Variance of $\hat{\beta}$
Distribution of $\hat{\beta}$ under the Classical Assumptions
Gauss Markov Theorem
BLUE
Maximum likelihood

Questions

1. Suppose we do a regression and obtain TSS = 1,231, SSR = 108 and N = 212. Compute the F statistic.

2. Consider a sample of observations x_i from a probability distribution with mean μ and variance of 1 (so $E(x_i) = \mu$ and $var(x_i) = 1$ for $i = 1,..,N$). Prove that the following are unbiased estimators of μ.
   (a) $\Sigma x_i / N$
   (b) $(x_1 + x_N)/2$
   (c) x_1

3. Which of the three estimators in question 2 is the best one, in the sense of having the minimum variance?

4. Do Chapter 3 question 3(a—c).

5. Suppose we want to estimate the least squares slope coefficient $\hat{b}$ in the linear model

$$Y_i = bX_i + e_i.$$

   in which the mean of X and the mean of Y are both zero. Which of the following is the correct equation?

   a) $\hat{b} = \sum_{i=1}^{N}(X_i - Y_i)^2 \,/\, \sum_{i=1}^{N} Y_i^2$

   b) $\hat{b} = \sum_{i=1}^{N}(X_i - \bar{X})^2 \,/\, \sum_{i=1}^{N}(Y_i - \bar{Y}){X_i}^2$

   c) $\hat{b} = \sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) \,/\, \sum_{i=1}^{N}(X_i - \bar{X})^2$

   d) $\hat{b} = \sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) \,/\, \sum_{i=1}^{N}(e_i - \bar{e})^2$

6. The “least squares” rule estimates the value of the slope coefficient β that minimizes the sum of the squared errors implied by the regression equation. Suppose we use a different rule that

estimates the value of β that maximizes the likelihood that we have the correct distribution function for our estimate of β. If we measure "likelihood" using the function shown on page 74, will we get the same estimate each way?

7. Consider the linear model $Y_i = a + bX_i + e_i$. Answer True or False:

____ The X_i values are assumed to be fixed and non-random.

____ The e_i values are assumed to be uncorrelated with the Y_i values.

____ The maximum likelihood estimate of the slope $\hat{b}$ equals the least squares estimate.

____ The least squares slope estimator $\hat{b}$ is a non-random variable.

____ If the Classical Assumptions hold, the expected value of $\hat{b}$ is b.

Unit 6: Confidence Intervals and Hypothesis Tests

Readings: Chapters 3.5—3.7

Topics
Manipulating inequalities
Confidence interval for β
Hypothesis testing
Modification when σ^2 is unknown

Questions

1. Suppose we run a regression of a variable Y on another variable X plus a constant and obtain $\hat{\beta} = 3.0$ and standard error $s_b = 0.31$.
   (i) If we assume s_b is the true value, what is the 95% confidence interval for β?
   (ii) Is $\hat{\beta}$ statistically significant at 5%?
   (iii) What value of s_b would imply $\hat{\beta}$ is not significant at 5%?

2. If the sample size is $N = 62$ and we treat s_b as an estimate, what is the 95% confidence interval for β?

3. If a regression model $Y_i = \beta X_i + \epsilon_i$ has Normally-distributed errors, the maximum likelihood estimator $\hat{\beta}_m$ can be derived by maximizing the log-likelihood function

$$l(\beta_m, \sigma^2) = \ln\left\{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\right\} - \frac{1}{2\sigma^2}\Sigma_{i=1}^{N}(Y_i - \beta_m X_i)^2$$

   with respect to β_m. Since σ^2 is also unknown we need to estimate it. The maximum likelihood estimator $\hat{\sigma}_m^2$ can be derived in the same way, by maximizing $l(\beta_m, \sigma^2)$ with respect to σ^2. The derivative is

$$\frac{\partial l}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\Sigma(Y_i - \beta_m X_i)^2.$$

   Prove that the maximum likelihood estimator $\hat{\sigma}_m^2$ is biased. Hint: set the derivative equal to zero and solve for σ^2. You may use the fact that the least squares estimator s^2 is unbiased, which in this case is written:

$$E[s^2] = E\left[\frac{1}{N-1}\Sigma\left(Y_i - \hat{Y}_i\right)^2\right] = \sigma^2.$$

4. In the above question, suppose we have a sample size of $N = 271$ observations. Is the maximum likelihood estimate of σ^2 biased too large or too small, and by how much?

5. Suppose we have a sample of N=200 observations on variables X_i and Y_i. We want to estimate a simple model $Y_i = \alpha + \beta X_i + \epsilon_i$ and we do it as follows. There are 199 lines connecting points

(X_1, Y_1) to (X_i, Y_i) where $i = 2, \dots, N$. We can use simple geometry to compute the slope of each such line as

$$b_i = \frac{Y_i - Y_1}{X_i - X_1}.$$

Suppose we estimate β by taking the average of those slopes. We can call this estimator $\hat{b}$. Prove that $\hat{b}$ is unbiased.

6. If we compare the variance of $\hat{b}$ in Question 5 to that of $\hat{\beta}$, the OLS estimator, will it have a smaller variance?

7. What happens to the confidence interval of β when we do not assume σ^2 is known, and instead treat it as an unknown parameter needing to be estimated?

8. Suppose we run a regression $Y_t = a_0 + a_1 X_t + a_2 W_t + e_t$ on 63 observations and the results are as follows.

   Sample size = 63
   Total Sum of Squares = 110293
   Sum of Squared Residuals = 23722

Coefficient	Estimate	Standard Deviation
$\hat{a}_0$	-0.532	0.711
$\hat{a}_1$	5.260	2.96
$\hat{a}_2$	-4.790	3.11

   (a) What is the 95% confidence interval for $\hat{a}_0$?
   (b) What is the 99% confidence interval for $\hat{a}_1$?
   (c) Using a 5% significance level, is $\hat{a}_2$ statistically significant?

9. If Z is standard normal we can write the probability equation $\Pr(-1.96 \le Z \le 1.96) = 0.95$. We form a z-score using the OLS coefficients using

$$Z = \frac{\hat{\beta} - \beta}{s_\beta}$$

   where s_β is the standard deviation of $\hat{\beta}$. Plug the above expression for Z into the probability equation to derive an expression for the 95% confidence interval for β.

10. Consider the variables X and Y where we assume the mean of each is zero. Denote β_{yx} as the slope coefficient from regressing Y on X and denote β_{xy} as the slope coefficient from regressing X on Y. Prove that $\frac{\beta_{yx}}{\beta_{xy}} = \frac{var(Y)}{var(X)}$.

Unit 7: Multiple Regression Estimation and Testing

Readings: Chapters 4.1—4.3, 4.4.1

Topics
Classical assumptions for multiple regression
Changes to basic formulas
Hypothesis testing
Choice of explanatory variables
Testing linear restrictions

Questions

1. Suppose you run a multiple regression and your results look as follows.

Coefficient	Estimate	t-Statistic
$\hat{a}_0$	0.508	0.711
$\hat{a}_1$	1.260	0.024
$\hat{a}_2$	0.881	0.011

   R-squared: 0.771
   Regression F statistic: 9.26, *p*-value 0.000004

   What evidence do you have that you might have a problem of multicollinearity?

2. Referring to the formulas on page 101 in the textbook, explain why a high level of correlation between two explanatory variables means the *t* statistics will likely be small.

3. Fill in the formulas in following chart.

	Simple regression	Multiple regression
s^2	$\frac{\Sigma\hat{\epsilon}_i^2}{N-2}$	
d.f.	$N-2$	
R^2	$1-\frac{\Sigma\hat{\epsilon}_i^2}{\Sigma(Y_i-\bar{Y})^2}$	
Regression F	$\frac{(N-2)R^2}{1-R^2}$	

4. Indicate True or False for the following statements.

   (a) The classical assumptions no longer apply in a multiple regression.
   (b) The regression line goes through the sample mean of all the explanatory variables.
   (c) Omitted variable bias does not arise if the omitted variable is uncorrelated with the included variables.
   (d) If irrelevant explanatory variables are included the other coefficients will be biased.
   (e) An F test can be used to test multiple coefficient restrictions at once.

5. Consider the linear regression model

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \epsilon_i$$

   Write out the restricted version of this model that could be used to test the following linear restrictions:

   (a) $\beta_1 = \beta_2 = 0$
   (b) $\beta_1 = \beta_2$
   (c) $\beta_1 + \beta_2 = 1$
   (d) $\beta_1 - \beta_2 = 1$
   (e) $\beta_1 = 0; \beta_2 + \beta_3 = 0$

6. Run the following program in R Studio

```
rm(list=ls(all=TRUE))
fruit=read.table("orange.csv", sep=",", skip=2)
organic = fruit$V1
regular = fruit$V2
new = log(organic) + sqrt(regular)
lm(organic ~ regular + new)
```

   Following the instructions in Section 8.1 of the R Studio Tutorial, run a linear hypothesis test on whether the coefficients on both *regular* and *new* equal zero.

   Was that information already available in the `summary()` output from the lm command?

7. Suppose we regress variable Y_i on variable X_i using the model

$$Y_i = b_0 + b_1 X_i + e_i.$$

   Prove that the regression line passes through the point $(\bar{Y}, \bar{X})$.
   HINT: start by writing out the formula for $\bar{Y}$.

8. Consider the linear regression model

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

where $\bar{Y} = \bar{X}_1 = \bar{X}_2 = 0$. The least squares formula for the variance of $\hat{\beta}_1$ is

$$var(\hat{\beta}_1) = \frac{\sigma^2}{(1 - r^2)\Sigma X_{1i}^2}$$

where σ^2 is the error variance and r is the correlation between X_1 and X_2.

With reference to this formula explain what is the problem of multicollinearity and how it affects the results of a multiple regression model.

9. Suppose the true model is

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

but we estimate

$$Y_i = \alpha + \beta_1 X_{1i} + \epsilon_i$$

instead. Despite the possibility of Omitted Variable Bias, under what circumstance will we nonetheless obtain an unbiased estimate of β_1?

10. State the null hypothesis that is tested using a test statistic that consists of the estimator $\hat{\beta}_1$ divided by its standard error. What distribution do we use to compute the critical values?

UNIT 8: FUNCTIONAL FORMS

Readings: Chapters 4.5, 2.3.6

Topics
Linearizing a model using logs
Quadratic forms
Interaction terms
Dummy variables and slope coefficients

Questions

1. Suppose you have data on labour (L_i), Capital (K_i) and output (Y_i). You want to estimate a Cobb-Douglas production function $Y = AK_i^{\alpha}L_i^{\beta}$ and test for constant returns to scale ($\alpha + \beta = 1$). Explain how you could do this using linear regression and an *F* test.

2. The Environmental Kuznets Curve hypothesis states that, at low average income levels, national pollution tends to rise as the economy grows, but then turns around at a certain point and declines as income grows further. In other words, pollution tends to follow an upside-down U shaped path when plotted against income. Suppose you have data on pollution emissions P_t and average real income Y_t for a country. Write down a model that would let you test if the relationship between pollution and income is linear or whether it follows the Environmental Kuznets Curve.

3. Using the data in the file wagedisc.xls, figure out a functional form and estimate a model that lets you find out if Males benefit (i.e. obtain higher salaries) from additional years of experience the same as, more than or less than females.

4. Use the wagedisc.xls data to estimate a model that will allow you to determine if additional years of experience have a diminishing marginal benefit in your sample.

5. Use the wagedisc.xls data to estimate a model that will allow you to determine if additional years of experience are of greater benefit to individuals with relatively higher education.

6. On page 103 of the text book the regression *F* statistic (testing $R^2 = 0$) is shown as

$$F = \frac{R^2}{1 - R^2}\frac{(N - K - 1)}{k}.$$

On page 104, an *F* test on a linear hypothesis involving the regression coefficients is expressed as

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(N - k - 1)}.$$

Suppose we test whether all the slope coefficients are equal to zero, or $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$. Show that these two versions of the *F* statistic will be identical.

UNIT 9: HETEROSKEDASTICITY

Readings: Chapters 5.1—5.3

Topics
Violation of the constant variance assumption
Testing for heteroskedasticity
GLS transform
Heteroskedasticity-robust variance estimation

1. Suppose you have a regression model $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$ where the error term is heteroskedastic and the variance is known to follow the process $V(e_i) = \sigma^2 Z_i$. Assuming you have observations on Z_i write out a transformed regression model that would have homoscedastic errors and show that the error variance is now constant.

2. Obtain the spreadsheet **educ.xls** from the CourseLink page and convert it to a CSV file. Read in the data and regress EDUC on GDP and POP.
   (a) Consult Section 8.2 of the R Studio Tutorial to see how to order data. Use this method to order your data according to GDP, then conduct a Goldfeld-Quandt test, leaving out the middle 20% of the sample (8 observations). Verify that the test score is 0.51 if you sort according to POP, but is 11.64 if you sort according to GDP. The p-value for $F_{12,12} = 11.64$ is 0.00008.
   (b) Conduct a Breusch-Pagan test and verify that it has a value of 5.57 (which has a p-value of 0.062).
   (c) Explain the challenge in deciding if the model has heteroskedasticity based on the tests you conducted.

3. Run a HCE estimator using White's standard errors. Are the coefficients on GDP and POP significant?

4. Answer the following questions True or False.

   ____________ If heteroskedasticity is present the OLS estimator will be biased.

   ____________ Homoskedastic errors violate one of the classical assumptions.

   ____________ The challenge with using GLS is that we might not know the true form of the heteroskdasticity.

   ____________ When using White's HCE standard errors we do not need to re-estimate the regression slope coefficients.

   ____________ Heteroskedasticity is likely to arise when using cross-sectional data.

UNIT 10: AUTOCORRELATION

Readings: Chapter 5.4

Topics
Time series notation
Violation of the no-covariance assumption
AR1 and AR(p) processes
Tests for AR1: LM test, Box-Pierce & Ljung
Durbin-Watson score
GLS estimation via Cochrane-Orcutt procedure
HAC estimators

Questions

1. Suppose the regression error e_t follows an AR1 process $e_t = \rho e_{t-1} + u_t$ where $u_t \sim N(0, \sigma^2)$. Derive the variance of e_t.

2. Suppose we have a regression equation $y_t = \alpha + \beta X_t + e_t$, and e_t follows an AR1 process $e_t = \rho e_t + u_t$ where $u_t \sim N(0, \sigma^2)$. Derive a transformed version of this model that has an error term that is not autocorrelated, which could be used to estimate the distribution of $\hat{\beta}$.

3. Obtain the data file **computer.xls** from the Courselink site and convert it to a CSV file. Write a program to read in the data, rename the variables 'ch.sales' and 'ch.comp.purch' (for, respectively, the percent change in monthly sales and percent change in computer purchases) and then regress sales on comp. What is the slope coefficient, and is it significant?

4. Using the steps in Section 8.4 of the R Studio Tutorial, perform a Breusch-Godfrey test for autocorrelation of order 1 and another test for autocorrelation of order 2. Is there evidence for AR1 or AR2 errors? Might there be even higher order autocorrelation?

5. Download and install the `orcutt` package in R. Run the Cochrane-Orcutt procedure on your regression from Question 3 (see Section 7.4 of the R manual for instructions). What effect does this procedure have on the standard error of the slope coefficient?

6. Since we know we have higher-order autocorrelation than the Cochrane-Orcutt procedure can handle we can try a HAC estimator. The Newey-West procedure is applicable for autocorrelation as well as heteroskedasticity. Apply the method in Section 8.4 of the R Studio Tutorial and report the Newey-West standard errors and HAC t-statistics.

7. Answer the following questions True or False.

   ____________ If autocorrelation is present the OLS estimator will be biased.

   ____________ Autocorrelated errors violate one of the classical assumptions.

   ____________ The Cochrane-Orcutt procedure is valid for any form of autocorrelation.

_____________ When using Newey-West standard errors we do not need to re-estimate the regression slope coefficients.

_____________ Autocorrelation is likely to arise when using time series data.

UNIT 11: ENDOGENEITY AND INSTRUMENTAL VARIABLES

Readings: Chapter 5.5

Topics
Endogenous versus exogenous variables
Consistency
Probability limits
Instrumental variable estimator
Measurement error in the explanatory variable
Indirect least squares

Questions

1. Suppose we have a simple regression model

$$Y_i = \beta X_i + e_{1i}.$$

But we also suspect that X_i is dependent on Y_i according to

$$X_i = \gamma Y_i + e_{2i}.$$

If both statements are true, prove that $cov(X_i, e_{1i}) \neq 0$.

2. Suppose a random variable X has a mean μ and variance σ^2. Prove $\text{plim}(a\bar{X}) = a\mu$ where a is a positive constant and $\bar{X}$ is the mean of sample size N.

3. Follow the steps in section 8.5 of the R Studio Tutorial to estimate the effect of TEST on the LogEARN variable, using SM as an instrument for TEST. Now add SF and SIBLINGS as additional instruments. How do the IV regression results change?

4. Suppose we have a simple regression model $Y_i = X_i\beta + e_i$ and we have an instrument Z_i for X_i. If we use the IV method we will get the formula $\hat{\beta}_{IV}$ as shown on page 153 in the textbook:

$$\hat{\beta}_{IV} = \frac{\Sigma Z_i Y_i}{\Sigma X_i Z_i}$$

Now suppose instead we regress X_i on Z_i using the equation $X_i = \gamma Z_i + v_i$ and obtain the fitted values $\hat{X}_i$. If we use these instead of Z_i as our instrument, will we get a different estimate of β?

5. Answer the following questions True or False.

_____________ If endogeneity is present the OLS estimator will be biased.

_____________ Random X variables violates one of the classical assumptions.

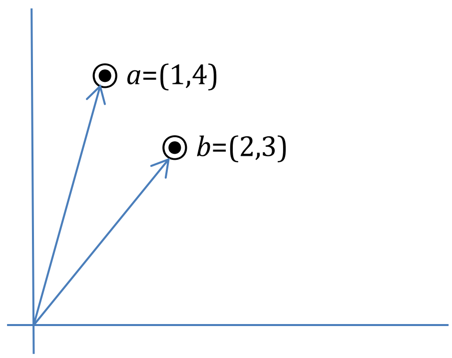
_____________ The challenge with using Instrumental Variables is that we might not be able to find a suitable instrument.

_____________ White's HCE standard errors also corrects the bias due to endogeneity.

_____________ Errors in measurement of the explanatory variables can give rise to the problem of random *X*'s.

UNIT 12: MATRIX REPRESENTATION OF OLS REGRESSION

Vectors and matrices
Consider two points in (X,Y) space;



The arrows, or vectors, can be summarized by the pair of coordinate points, with the first number representing the horizontal distance and the second number representing the vertical distance.

Vectors are written in column form, as in

$$a = \begin{bmatrix} 1 \\ 4 \end{bmatrix} ; b = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

If the angle between two vectors is θ, then the cosine of θ indicates how "parallel" they are on a scale from 0 to 1. If they are exactly parallel then the angle between them is zero, and $\cos(0) = 1$. If they are at right angles (or *orthogonal*) the angle between them is 90 degrees, and $\cos(90) = 0$.

If we add the vectors we get $c = a + b = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$. This is the vector whose end point can be pictured by shifting b up along a so it starts at the tip of a. Then the arrow tip of b would point to c.

The length of a vector is computed by squaring each element, adding them up, then taking the square root of the result. We write this as $||a|| = \sqrt{1^2 + 4^2} = \sqrt{17}$. Note that $||a||$ means the length of a.

The operation of multiplying vectors is done using a so-called *dot* (or *scalar*) *product*. It has a completely different intuition than multiplying numbers. Unfortunately it isn't easy to provide the intuition since the main examples come from physics. If we think of vectors as forces in space, the dot product of two vectors is the projection of one vector into the direction of the other, scaled by the length of the second vector. Formally this is done using an expression involving cosines and the vector lengths. (So, not very intuitive…)

Fortunately there is a simpler formula. For two vectors g and h, each with two elements, the dot product is:

$$g \cdot h = (g_1 h_1 + g_2 h_2).$$

In other words, the dot product of two vectors is just the product of each of the elements in sequence added up. So using our earlier example, $a \cdot b = 2 + 12 = 14$.

Note that the dot product of any vector with itself is its length. Vectors that are at right angles to each other are called *orthogonal*. They have a dot product of zero, since the projection of one onto the other collapses on a single point and has no length. You can see an example of this if you graph the vectors (1,1) and (2, -2). They are at right angles and their dot product is $2 - 2 = 0$. When we start to represent variables in a data set as vectors, orthogonality will correspond to the vectors being completely *uncorrelated* with each other. So you can also think of a dot product as related to correlation.

Matrices

If we group these vectors together in a row, as in $[a\ b]$ we have a matrix. Call it C:

$$C = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}.$$

Notation: Each row is called a "row"
Each column is called a "column"
Each number is called an "element"

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

c_{21} is the element in the 2nd row, 1st column

A matrix element has 2 subscripts, denoting the row and column in that order. The matrix has dimensions *rc* where *r*= the number of rows and *c*= the number of columns. So if we write C_{22} it indicates the dimensions of C.

Simple operations: addition, scalar multiplication, transposing

Consider 2 matrices $A_{Nk} = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{Nk} \end{pmatrix}$ and $B_{Nk} = \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{N1} & \cdots & b_{Nk} \end{pmatrix}$

Addition is easy: just add each of the corresponding elements:

$$C_{Nk} = A_{Nk} + B_{Nk} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1k} + b_{1k} \\ \vdots & \ddots & \vdots \\ a_{N1} b_{N1} & \cdots & a_{Nk} + b_{Nk} \end{pmatrix}$$

Note that the matrices being added must have the same dimensions.

Scalar multiplication is also easy. A scalar is just a single number, as in $z = 3$. When multiplying a scalar times a matrix, just multiply each element by the scalar.

$$z \times A_{Nk} = \begin{pmatrix} z \times a_{11} & \cdots & z \times a_{1k} \\ \vdots & \ddots & \vdots \\ z \times a_{N1} & \cdots & z \times a_{Nk} \end{pmatrix}.$$

Transposing means switching rows and columns; turning column 1 into row 1, column 2 into row 2, etc.

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 5 & 6 \end{bmatrix}.$$

The transpose is denoted A^T and it is

$$A^T = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 1 & 6 \end{bmatrix}.$$

Note that if the dimensions of A are $N \times k$ then the dimensions of A^T are $k \times N$.

Multiplication

Suppose we have $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 5 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 9 \\ 1 & 3 \\ 0 & 4 \end{bmatrix}$. What is $A \times B$?

The "natural" answer would be to multiply the elements, as in

$$A \times B = \begin{bmatrix} 1 \times 3 & 3 \times 9 \\ 2 \times 1 & 1 \times 3 \\ 5 \times 0 & 6 \times 4 \end{bmatrix}.$$

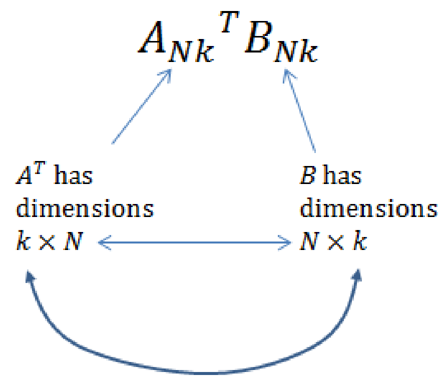
But this is incorrect. The reason is that the columns are vectors, and we want to get a matrix of dot products, not scalar products. As we saw before, vector dot products involve multiplying each element and adding them up.

$$x = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, y = \begin{bmatrix} 5 \\ 2 \\ 3 \end{bmatrix} \text{ so } x \cdot y = (1 \times 5) + (3 \times 2) + (2 \times 3) = 17.$$

A and B both have 2 columns. So the product has to have 4 dot products. The convention for writing this out is to conform the dimensions so that the *first* matrix supplies vectors in rows and the *second* matrix supplies vectors in columns, and we form dot products between each of the rows in the first matrix and each of the columns in the second matrix. In this example it requires transposing A and multiplying it into B.

$$A^T B = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} 3 & 9 \\ 1 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 35 \\ 10 & 54 \end{bmatrix}.$$

Note what happens to the dimensions. For the matrices to be *conformable* requires that the inner, adjacent numbers be the same, and the outer numbers determine the dimensions of the product, i.e. *kNNk*:



The product $C = A^T B$ has dimensions $k \times k$.

A matrix times a vector

This one's easy. If the rows of the matrix represent the vectors it will contribute to the dot products, multiply each row by the vector. For example, suppose X has N rows and k columns and b has k rows and 1 column. Then the product is as follows.

$$Xb = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}$$

$$= \begin{bmatrix} x_{11}b_1 + x_{12}b_2 + \cdots + x_{1k}b_k \\ \vdots \\ x_{N1}b_1 + x_{N2}b_2 + \cdots + x_{Nk}b_k \end{bmatrix}.$$

The inverse of a matrix

The inverse of a number is just the reciprocal, i.e. the number which, multiplied by the original, yields 1. For instance, the inverse of 5 is $\frac{1}{5}$, since $5 \times \frac{1}{5} = 1$. We write inverse using the "-1" superscript: $5^{-1} = \frac{1}{5}$.

If A is a matrix, what is A^{-1}? It is the matrix that, when multiplied by A yields the "identity" matrix, which is a matrix of zeroes except for the diagonal positions which are ones. For instance, a 3-dimensional identity matrix is

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that only a *square* matrix can have an inverse. A square matrix has the same number of rows and columns. Its inverse will then have the same dimension.

How to compute the inverse of a matrix.

Step 1: Buy a computer.
Step 2: Get the computer to invert your matrix.

Computing the inverse of a matrix is very laborious. But computers seem to enjoy it and are very fast.

Not all square matrices have an inverse. For the inverse to exist the matrix must be of *full rank*: that means all its rows and columns must be independent. It must not be possible to express any one row (or column) as a linear combination of any other row (or column).

Regression model in matrix notation

Suppose our data set runs from $t = 1, \dots, T$. We can write the k-variable multiple regression model as follows:

$$Y_t = b_0 + b_1 X_{1t} + b_2 X_{2t} + \cdots + b_k X_{kt} + e_t.$$

Now define a few matrices.

$$X = \begin{bmatrix} 1 & x_{11} & \cdots x_{k1} \\ \vdots & \vdots & \cdots \vdots \\ 1 & x_{1T} & \cdots x_{kT} \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}$$

$$b = \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix}$$

$$e = \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix}$$

We can write the regression model as

$$y = Xb + e.$$

You should be able to verify that this is an equivalent representation.

The sum of the squared residuals is now written $e^T e$.

We want to minimize this by choice of the vector b. The solution is our OLS result

$$\hat{b} = (X^T X)^{-1} X^T y.$$

You should be able to verify that this is a $(k + 1) \times 1$ vector. The variance-covariance matrix of $\hat{b}$ is

$$V(\hat{b}) = \sigma^2 (X^T X)^{-1}.$$

Note this is a $(k+1) \times (k+1)$ matrix. The variance of first coefficient (b_0) is the first element down the diagonal, and the variances of the others are the remaining diagonal elements in order. The off-diagonal elements are the covariances between the slope coefficients.

Questions

1. Consider two vectors $a = \begin{bmatrix} 4 \\ 2 \\ -3 \end{bmatrix}$ and $b = \begin{bmatrix} -2 \\ 7 \\ 2 \end{bmatrix}$. Show that they are orthogonal to each other.

2. If $A = \begin{bmatrix} 9 & -1 \\ -3 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -4 \\ 12 & 1 \end{bmatrix}$ compute $C = AB$.

3. What is $A^T B$?

4. If a matrix A has dimensions (100, 13) and a matrix B has dimensions (13, 12), what are the dimensions of the product AB? What are the dimensions of the matrix $A^T A$? What would be the dimensions of $(A^T A)^{-1} A^T$?

5. Write the regression equation

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 X_{3i} + e_i$$

in matrix notation, assuming that the sample length is $i = 1, \dots, N$.

6. Verify that the OLS estimator $(X^T X)^{-1} X^T Y$ has the same dimensions as $\hat{b}$ (assume that $\hat{b}$ has k elements).

7. Suppose a linear regression yields the following results:

```
Call:
lm(formula = Y ~ X1 + X2 + X3)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.047  -6.539  -1.905   3.603  91.812 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -21.9965    10.6005  -2.075   0.0385 *  
X1            0.1514     0.2481   0.610   0.5420    
X2            2.3443     0.2167  10.820  < 2e-16 ***
X3            7.0064     1.1065   6.332 5.13e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.85 on 536 degrees of freedom
Multiple R-squared:  0.2294,    Adjusted R-squared:  0.2251 
F-statistic: 53.19 on 3 and 536 DF,  p-value: < 2.2e-16
```

The associated variance-covariance matrix of the coefficients is

```
            (Intercept)            X1            X2           X3
(Intercept) 112.3712844 -2.5188270837 -0.6747417173 -0.722341298
X1           -2.5188271  0.0615728550  0.0008563438  0.004031909
X2           -0.6747417  0.0008563438  0.0469471794 -0.003942115
X3           -0.7223413  0.0040319092 -0.0039421146  1.224399585
```

Where do the square roots of the diagonal elements of the above matrix appear in the regression results?

APPENDIX: WHY WE USE (*N*-1) IN THE VARIANCE FORMULA

Suppose we have a variable X and a sample of N observations x_i where $i = 1, \dots, N$.

The population mean and variance are defined as:

$$\mu = E(X)$$

and

$$var(X) = \sigma^2 = E(X - \mu)^2$$

Since we usually don't know μ and σ^2 we have to estimate them using a data sample. When the x_i's are independent, if we know μ we can write

$$E(X - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2.$$

Suppose we use $\bar{X}$ to estimate μ. If $\bar{X}$ denotes a sample mean, then note that

$$var(\bar{X}) = E(\bar{X} - \mu)^2$$

and

$$var(\bar{X}) = var\left(\frac{1}{N} \times \Sigma x_i\right) = \frac{1}{N} \times N^2\sigma^2 = \frac{1}{N}\sigma^2.$$

Since we don't know σ^2 we have to estimate it. We want an estimator whose expected value is the same as the population parameter. Suppose we use the definition of population variance to get:

$$s_x^2 = \frac{1}{N}\Sigma(x_i - \bar{X})^2$$

The subscript *x* means "wrong" – as I will show. The bias in the estimate is $E(\sigma^2 - s_x^2)$. This should equal zero. Expanding it out:

$$E(\sigma^2 - s_x^2) = E\left(\frac{1}{N}\Sigma(x_i - \mu)^2 - \frac{1}{N}\Sigma(x_i - \bar{X})^2\right)$$

$$= \frac{1}{N}E\big(\Sigma(x_i^2 - 2x_i\mu + \mu^2) - \Sigma(x_i^2 - 2x_i\bar{X} + \bar{X}^2\big)$$

$$= \frac{1}{N}E\big(\Sigma(x_i^2 - x_i^2 - \bar{X}^2 + \mu^2 - 2x_i\mu + 2x_i\bar{X}\big)$$

$$= \frac{1}{N}E(N\mu^2 - N\bar{X}^2 + 2\Sigma x_i(\bar{X} - \mu))$$

$$= E(\mu^2 - \bar{X}^2 + 2\bar{X}(\bar{X} - \mu))$$

$$= E(\mu^2 - \bar{X}^2 + 2\bar{X}^2 - 2\bar{X}\mu)$$

$$= E(\mu^2 - 2\bar{X}\mu + \bar{X}^2)$$

$$= E(\bar{X} - \mu)^2$$

$$= var(\bar{X}) = \frac{\sigma^2}{N}.$$

So we have $E(\sigma^2 - s_x^2) = \sigma^2/N$. This rearranges to

$$E(s_x^2) = \frac{N-1}{N}\sigma^2$$

So s_x^2 is biased small. To correct this, we can get an unbiased estimator using

$$\frac{N}{N-1}s_x^2 = \frac{1}{N-1}\Sigma(x_i - \bar{X})^2$$

ANSWERS

Unit 1.

2. $(0.16 \times 0) + (0.24 \times 1) + (0.23 \times 2) + (0.18 \times 3) + (0.09 \times 4) + (0.06 \times 5) + (0.04 \times 6) = 2.14.$

3. $(0.16 \times -2.14^2) + (0.24 \times -1.14^2) + (0.23 \times -0.14^2) + (0.18 \times 0.86^2) + (0.09 \times 1.86^2) + (0.06 \times 2.86^2) + (0.04 \times 3.86^2) = 2.5804$

4. $(0.16 \times 0) + (0.24 \times 1) + (0.23 \times 4) + (0.18 \times 9) + (0.09 \times 16) + (0.06 \times 25) + (0.04 \times 36) - 2.14^2 = 2.5804$

8. 20 + 20 = 40.
9. 7 as well.
10. $var(Y) = 2^2 \times 1 = 4.$

11. $var(9X + 2Y) = (9^2 \times 4) + (2^2 \times 3) + 2 \times (9 \times 2) \times 3.6 = 324 + 12 + 129.6 = 465.6.$

Unit 2.

1. Since Y_i is a constant, $Y_i = \bar{Y}$ so $Y_i - \bar{Y} = 0$ for all i. Then $E(Y_i - \bar{Y})(X_i - \bar{X}) = 0 \times E(X_i - \bar{X}) = 0.$

2. Squares refers to the sum of the squared residuals. We are finding the coefficients that give us the minimum value of this, or the "least" value.

3. We have N observations on X_i and we want to approximate them using a single number $\bar{X}$. The error terms associated with each observation will be $e_i = (X_i - \bar{X})$. The sum of the squared errors E will be $E = \Sigma_{i=1}^{N}(X_i - \bar{X})^2$. To minimize this expression take the first derivative with respect to $\bar{X}$, set it equal to zero and solve for $\bar{X}$:

$$\frac{\partial E}{\partial \bar{X}} = -2\left(\Sigma_{i=1}^{N}(X_i - \bar{X})\right) = 0$$

$$\Sigma_{i=1}^{N} X_i = N\,\bar{X}$$

$$\frac{1}{N}\Sigma_{i=1}^{N} X_i = \bar{X}.$$

4. Note $R^2 = 1 - \frac{SSR}{TSS}$. If we use calculus to choose a value of $\hat{\alpha}$ that minimizes $\sum \epsilon_i^2$ we obtain the usual formula for the mean:

$$\hat{\alpha} = \frac{1}{N}\sum Y_i = \bar{Y}.$$

So the regression model yields

$$Y_i = \bar{Y} + \epsilon_i.$$

This implies the sum of squared residuals is

$$SSR = \sum (Y_i - \bar{Y})^2.$$

But by definition the total sum of squares is

$$TSS = \sum (Y_i - \bar{Y})^2.$$

So $R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum(Y_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - 1 = 0.$

5. $var(Y_i) = 4 \times 9.1 = 36.4.$
   Using the formula for correlation and the fact that $Y_i = 2X_i$:

$$r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\Sigma(X_i - \bar{X})^2}\sqrt{(\Sigma(Y_i - \bar{Y})^2}}$$

$$= \frac{\Sigma(X_i - \bar{X})(2X_i - 2\bar{X})}{\sqrt{(\Sigma(X_i - \bar{X})^2}\sqrt{(\Sigma(2X_i - 2\bar{X})^2}}$$

$$= \frac{2\Sigma(X_i - \bar{X})(X_i - \bar{X})}{\sqrt{(\Sigma(X_i - \bar{X})^2}\sqrt{4(\Sigma(X_i - \bar{X})^2}}$$

$$= \frac{\Sigma(X_i - \bar{X})^2}{\sqrt{(\Sigma(X_i - \bar{X})^2}\sqrt{(\Sigma(X_i - \bar{X})^2}}$$

$$= \frac{\Sigma(X_i - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}$$

$$= 1.$$

6. The result should be

```
Call:
lm(formula = regular ~ new)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8842 -3.9454 -0.7296  3.5742 10.9187

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -202.7445     2.2723  -89.22   <2e-16 ***
new           20.5767     0.1335  154.15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.475 on 179 degrees of freedom
Multiple R-squared:  0.9925,    Adjusted R-squared:  0.9925
F-statistic: 2.376e+04 on 1 and 179 DF,  p-value: < 2.2e-16
```

The slope coefficient is 20.5767.
The t statistic is 20.5767 / 0.1335 = 154.13. (The difference is due to rounding).

7. $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$.

8. The residual.

9. a.

10. a) $\Sigma_{i=1}^{N}(W_i - \overline{W}^2)$

    b) $\Sigma_{i=1}^{N}(\widehat{W}_i - \overline{W}^2)$

    c) $\Sigma_{i=1}^{N}\hat{\epsilon}_i^2 \;\; or \;\; \Sigma_{i=1}^{N}(W_i - \widehat{W}_i^2)$

    d) $\frac{RSS}{TSS} = \frac{\Sigma_{i=1}^{N}(\widehat{W}_i - \overline{W}^2)}{\Sigma_{i=1}^{N}(W_i - \overline{W}^2)}$

    e) $\frac{(N-2)R^2}{1-R^2}$

Unit 3.

1. The following code will work.

```
##
## Koop Ch 2 Question 1
##
rm(list=ls(all=TRUE))

## Note there are 2 header rows
dat = read.csv(file="electric.csv", skip=2, header=FALSE)

## Rename variables
```

```
Y = dat$V1
output = dat$V2
pL = dat$V3
pK = dat$V4
pF = dat$V5

## Regression model
reg = lm(Y ~ output+pL+pK+pF)
summary(reg)
```

a) The output is:

```
Call:
lm(formula = Y ~ output + pL + pK + pF)

Residuals:
    Min      1Q  Median      3Q     Max 
-56.356  -4.899  -1.465   3.951  81.417 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept) -70.495109  12.695008  -5.553 1.76e-07 ***
output        4.731137   0.109451  43.226  < 2e-16 ***
pL            0.003627   0.001055   3.437 0.000814 ***
pK            0.280083   0.129488   2.163 0.032557 *  
pF            0.783460   0.165789   4.726 6.39e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.63 on 118 degrees of freedom
Multiple R-squared:  0.9438,    Adjusted R-squared:  0.9419 
F-statistic: 495.4 on 4 and 118 DF,  p-value: < 2.2e-16
```

Each of the coefficients is significant at 5% ($p < 0.05$) or less. As seen previously, output has a strongly positive effect on production costs. The price of Fuel (pF) does as well. Labour and capital costs also significantly increase production costs.

b) In this case the coefficient on output changes very little after the introduction of the price variables: it falls from 4.79 to 4.73. This implies that the price variables are likely not strongly correlated with the output variable.

d) Since all the coefficients are significant it would not make sense to drop any of them.

2. From the definition of RSS (p. 36), $RSS = \Sigma(\hat{Y}_i - \bar{Y})^2 = \Sigma(\hat{Y}_i)^2$ (since the mean of Y is zero). But since $\hat{Y}_i = \hat{\beta}X_i = 0$ (since the slope coefficient is zero) we have $RSS = 0$. Since $R^2 = RSS/TSS$ it must mean that $R^2 = 0$ too.

3. (a) 38; -79.45. (b) 0.458 (c) no.

4. The F statistic is 125 and its p-value is very close to zero. It means we can reject the null hypothesis that $R^2 = 0$.

Unit 4.

1. $E(Z) = \left(\frac{1}{\sigma}\right)(E(Y) - \mu) = \left(\frac{1}{\sigma}\right)(\mu - \mu) = 0.$
   $var(Z) = \left(\frac{1}{\sigma^2}\right)var(Y) = \left(\frac{1}{\sigma^2}\right)\sigma^2 = 1.$

2. (a) $\Pr(Z \geq 0) = 0.5$
   (b) $\Pr(Z \geq 1) = 1 - 0.5 - 0.3413 = 0.1587$
   (c) $\Pr(Z < 1) = 0.5 + 0.3413 = 0.8413$
   (d) $\Pr(Z \leq 1.5) = 0.5 + 0.1915 = 0.6915$
   (e) $\Pr(-0.5 \leq Z \leq 0.5) = 0.1915 + 0.1915 = 0.383$
   (f) $\Pr(-0.64 \leq Z \leq 0.83) = 0.2389 + 0.2967 = 0.5356$

3. The error terms are not autocorrelated, or in other words they are independent of each other across the sample.

4. The first one, $E(Y_i) = \beta X_i$ or $E(\epsilon_i) = 0$.

5. No, it does not mean that every observation lies on the regression line, only that on average the observations will lie on the regression line.

6. b). These assumptions determine how $\hat{\beta}$ is distributed.

7. The mean μ and the variance σ^2.

8. $s^2 = \frac{\Sigma(X_i - \bar{X}^2)}{N-1}$

9. Note that $t^2 = \frac{Z^2}{X/n}$. The numerator follows a χ^2 distribution with 1 degree of freedom. The denominator follows a χ^2 distribution with *n* degrees of freedom. This corresponds with the structure of a random variable that follows an *F* distribution with 1 and *n* degrees of freedom.

Unit 5.

1 $R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{108}{1231} = 0.912$. Then $F = \frac{(N-2)R^2}{1-R^2} = \frac{210 \times 0.912}{1-0.912} = 2176.4.$

2. (a) $E\left(\frac{\Sigma x_i}{N}\right) = \frac{NE(x_i)}{N} = \left(\frac{N\mu}{N}\right) = \mu$
   (b) $E(x_1 + x_N)/2) = (\mu + \mu)/2 = \mu$
   (c) $E(x_1) = \mu$

3. The variances are: (a) σ^2/N, (b) $\sigma^2/2$, (c) σ^2. As long as $N > 2$ (a) is the best.

4. (a) Yes:

$$\tilde{\beta} = \frac{\Sigma Y_i}{\Sigma X_i} = \frac{\Sigma(\beta X_i + e_i)}{\Sigma X_i} = \beta\frac{\Sigma X_i}{\Sigma X_i} + \beta\frac{\Sigma e_i}{\Sigma X_i} = \beta + \beta\frac{\Sigma e_i}{\Sigma X_i}$$

Take the expected value. The second term disappears and we are left with $E(\tilde{\beta}) = \beta$.

(b) $var(\tilde{\beta}) = \left(\frac{1}{\Sigma X_i}\right)^2 \Sigma var(X_i\beta + \epsilon_i) = \left(\frac{1}{\Sigma X_i}\right)^2 N\sigma^2$.

(c) The variance of $\hat{\beta}$ is $\frac{\sigma^2}{\Sigma X_i^2}$. The variance of $\tilde{\beta}$ can be written $\frac{N}{(\Sigma X_i)^2}\sigma^2$. Use the result shown in the text to show that this is larger than the variance of $\hat{\beta}$.

5. c) is correct. Set the means to zero and compare to the formula on page 67.

6. Yes. Maximizing the likelihood function requires us to minimize the sum of the squared error terms, so the solutions are the same.

7. True; False (the Y values are a linear function of the error terms so they must be correlated); True; False (it is a function of the Y's so it must also be a random variable); True.

Unit 6.

1. (i) $3.0 \pm 1.96 \times 0.31 = (2.39, 3.61)$.
   (ii) Since 0 is outside the confidence interval, the answer is yes.
   (iii) Solve $3.0 - 1.96 \times s_b = 0$ to get $s_b = 1.53$.

2. There are $62 - 2 = 120$ degrees of freedom. Use the t tables to get the 0.025 critical value for 60 degrees of freedom, which is 2.0. So the confidence interval is $3.0 \pm 2 \times 0.31 = (2.38, 3.62)$

3. Set the derivative to zero and rearrange it to get

$$\frac{1}{N}\Sigma(Y_i - \beta_m X_i)^2 = \hat{\sigma}_m^2.$$

   Since this does not correspond to s^2 we know that its expected value will not be σ^2, so it is biased.

4. The unbiased estimator of σ^2 is $\frac{1}{N-1}\Sigma(Y_i - \beta X_i)^2$. The two formulas are the same except for the fraction in front. Note that

$$\frac{1}{N} < \frac{1}{N-1}.$$

   So the maximum likelihood estimator is a bit too small (it underestimates the variance). The fraction by which is does so is $\frac{N-1}{N}$. For N=271, this corresponds to 270/271 = 0.9963. So the maximum likelihood variance estimate is too small by 1 – 0.9963 = 0.37%.

5. The formula for $\hat{b}$ is

$$\hat{b} = \left(\frac{1}{N-1}\right)\sum_{i=2}^{N}\frac{Y_i - Y_1}{X_i - X_1}.$$

Take the expected value, using $E(Y_i) = X_i\beta$ for all *i* (including *i*=1):

$$E[\hat{b}] = \left(\frac{1}{N-1}\right)\sum_{i=2}^{N}\frac{E[Y_i]-E[Y_1]}{X_i - X_1}$$

$$= \left(\frac{1}{N-1}\right)\sum_{i=2}^{N}\frac{\beta X_i - \beta X_1}{X_i - X_1}$$

$$= \left(\frac{1}{N-1}\right)\beta\sum_{i=2}^{N}\frac{X_i - X_1}{X_i - X_1}$$

$$= \left(\frac{N-1}{N-1}\right)\beta$$

$$= \beta.$$

6. No. It is a linear, unbiased estimator, so by the Gauss-Markov theorem the OLS estimator has the smallest variance.

7. It gets wider, unless the sample size is very large.

8. (a) With 60 degrees of freedom the 0.025 critical value of the *t* distribution is 2.000. So the 95% CI is $-0.532 \pm 2 \times 0.711 = (-1.954, 0.890)$.
(b) The 0.005 critical value of t_{60} is 2.660. So the 99% CI is $5.260 \pm 2.660 \times 2.96 = (-2.6136, 13.1336)$.
(c) Test $H_0{:}\, a_2 = 0$. The *Z* score is $-4.79/3.11 = -1.54$. Its absolute value is 1.54. The 0.025 critical value for t_{60} is 2.000. Since *Z*<2.000 we do not reject the null hypothesis, implying that $\hat{a}_2$ is not significant at 5%.

9. Follow the steps on page 76 of the textbook to get $\Pr\left[\hat{\beta} - 1.96s_\beta \le \beta \le \hat{\beta} + 1.96s_\beta\right] = 0.95$.

10. $\frac{\beta_{yx}}{\beta_{xy}} = \frac{\Sigma X_iY_i}{\Sigma X_i^2} \div \frac{\Sigma X_iY_i}{\Sigma Y_i^2} = \frac{\Sigma X_iY_i}{\Sigma X_i^2} \times \frac{\Sigma Y_i^2}{\Sigma X_iY_i} = \frac{\Sigma Y_i^2}{\Sigma X_i^2}$. Divide the top and the bottom by $N-1$ to get the result.

Unit 7

1. The *t* statistics are very low yet the *F* statistic is high and significant.

2. If r^2 is close to 1 then $(1-r^2)$ is close to zero, which makes the variance rise to a very high number. As a result the *t* stats will become very small.

3.

	Simple regression	Multiple regression
s^2	$\frac{\Sigma\hat{\epsilon}_i^2}{N-2}$	$\frac{\Sigma\hat{\epsilon}_i^2}{N-k-1}$

d.f.	$N-2$	$N-k-1$
R^2	$1-\frac{\Sigma\hat{\epsilon}_i^2}{\Sigma(Y_i-\bar{Y})^2}$	$1-\frac{\Sigma\hat{\epsilon}_i^2}{\Sigma(Y_i-\bar{Y})^2}$
Regression F	$\frac{(N-2)R^2}{1-R^2}$	$\frac{(N-k-1)}{k}\left(\frac{R^2}{1-R^2}\right)$

4. F, T, T, F, T

5. (a) $Y_i = \alpha + \beta_3 X_{3i} + \epsilon_i$
   (b) $Y_i = \alpha + \beta_1(X_{1i} + X_{2i}) + \beta_3 X_{3i} + \epsilon_i$
   (c) $Y_i - X_{1i} = \alpha + \beta_2(X_{2i} - X_{1i}) + \beta_3 X_{3i} + \epsilon_i$
   (d) $Y_i - X_{1i} = \alpha + \beta_2(X_{2i} + X_{1i}) + \beta_3 X_{3i} + \epsilon_i$
   (e) $Y_i = \alpha + \beta_3(X_{3i} - X_{2i}) + \epsilon_i$

6. Yes, this is what the regression F test shows.

7. The fitted regression line is $Y_i = \hat{b}_0 + \hat{b}_1 X_i + \hat{e}_i$. Then
$$\bar{Y} = \frac{1}{N}\Sigma Y_i = \frac{1}{N}\Sigma\left(\hat{b}_0 + \hat{b}_1 X_i + \hat{e}_i\right) = \frac{N}{N}\hat{b}_0 + \hat{b}_1\left(\frac{1}{N}\right)\Sigma X_i + \frac{1}{N}\Sigma\hat{e}_i = \hat{b}_0 + \hat{b}_1\bar{X}.$$
(The last step uses the fact that the residuals have a zero mean.)

8. The more correlated the variables are, the closer r^2 gets to 1. As r^2 approaches 1, $var(\hat{\beta})$ goes to ∞. However the formula for R^2 isn't affected. So we would get very high coefficient variances (and very small t statistics) yet the fit of the regression wouldn't go down.

9. If X_1 and X_2 are uncorrelated then there will be no Omitted Variable Bias.

10. $H_0: \beta_1 = 0$. Use the t tables.

Unit 8.

1. Take logs of both sides of the production function to yield $\ln(Y_t) = A + \alpha\ln(K_t) + \beta ln(L_t) + \epsilon_t$. Estimate this equation and call the sum of squared residuals USSR. Apply the linear restriction $\alpha + \beta = 1$ to get the restricted regression $(\ln(Y_t) - 1) = A + \beta(ln(L_t) - ln(K_t)) + \epsilon_t$. Estimate this equation and call the sum of squared residuals RSSR. Test whether the restriction significantly changes the sum of squared residuals using $F_{1,N-3} = (N-3) \times (RSSR - USSR)/USSR$.

2. Estimate $P_t = \alpha + \beta_1 Y_t + \beta_2 Y_t^2 + \epsilon_t$ and test if $\beta_2 \leq 0$.

3. Form a variable mEXP = MALE*EXP and estimate the regression model with explanatory variables MALE, EDUC, EXP and mEXP. The results should be:

```
Call:
lm(formula = Salary ~ MALE + EDUC + EXP + mEXP)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1989 -0.7114 -0.0294  0.7744  2.4752

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.01089    0.15111 132.429  < 2e-16 ***
MALE        -0.08004    0.29195  -0.274   0.7846
EDUC         0.50779    0.02788  18.213  < 2e-16 ***
EXP          1.38736    0.28578   4.855 4.73e-06 ***
mEXP        -0.64422    0.29097  -2.214   0.0292 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9843 on 95 degrees of freedom
Multiple R-squared:  0.869,     Adjusted R-squared:  0.8634
F-statistic: 157.5 on 4 and 95 DF,  p-value: < 2.2e-16
```

Since the coefficient on mEXP is negative and significant (*p*=0.0292) we conclude that additional years of experience are significantly less beneficial for men than for women.

4. If you add EDUC^2 to the model with MALE, EXP and EDUC the results will be

```
Call:
lm(formula = Salary ~ MALE + EXP + EDUC + EDUC2)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2289 -0.6596 -0.0302  0.7130  2.3841

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.040877   0.162683 123.189  < 2e-16 ***
MALE        -0.305803   0.286621  -1.067    0.289
EXP          0.771066   0.056014  13.766  < 2e-16 ***
EDUC         0.573981   0.066557   8.624 1.44e-13 ***
EDUC2       -0.005098   0.004424  -1.152    0.252
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.002 on 95 degrees of freedom
Multiple R-squared:  0.8641,    Adjusted R-squared:  0.8584
F-statistic:   151 on 4 and 95 DF,  p-value: < 2.2e-16
```

Since the coefficient on EDUC^2 is insignificant we do not reject the hypothesis that the effect is linear (rather than diminishing at the margin).

5. Adding an interaction term (EDEX = EDUC*EXP) yields

```
Call:
lm(formula = Salary ~ MALE + EXP + EDUC + EDEX)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2825 -0.7669  0.0045  0.7693  2.3823

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.09452    0.15067 133.365   <2e-16 ***
MALE        -0.21985    0.29179  -0.753    0.453
EXP          0.78506    0.06083  12.906   <2e-16 ***
EDUC         0.51710    0.03255  15.885   <2e-16 ***
EDEX        -0.01206    0.01524  -0.791    0.431
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.006 on 95 degrees of freedom
Multiple R-squared:  0.8631,    Adjusted R-squared:  0.8573
F-statistic: 149.7 on 4 and 95 DF,  p-value: < 2.2e-16
```

The interaction term is insignificant, so the marginal effect of experience is not influenced by the amount of education.

6. The restricted model involves regressing Y_i on a constant: $Y_i = \alpha + \epsilon_i$. This will yield an R^2 of zero (see Unit 2, question 3). Also, the number of restrictions in this case is k, the number of slope coefficients. Substituting $R_R^2 = 0$ and $q = k$ into the formula on page 104 yields

$$F = \frac{R_{UR}^2/k}{(1 - R_{UR}^2)/(N - k - 1)}.$$

This can be rearranged into

$$F = \frac{R_{UR}^2}{1 - R_{UR}^2}\frac{(N - K - 1)}{k}.$$

Note that R_{UR}^2 corresponds to R^2 in the formula on page 103, namely R^2 in the original model. So the two equations are identical.

Unit 9.

1. Divide through by $\sqrt{Z_i}$ to obtain

$$\frac{Y_i}{\sqrt{Z_i}} = \alpha\left(\frac{1}{\sqrt{Z_i}}\right) + \beta_1\left(\frac{X_1}{\sqrt{Z_i}}\right) + \beta_2\left(\frac{X_2}{\sqrt{Z_i}}\right) + \frac{e_i}{\sqrt{Z_i}}.$$

Now $V\left(\frac{e_i}{\sqrt{Z_i}}\right) = \left(\frac{\sigma^2 Z_i}{Z_i}\right) = \sigma^2$.

2—4. Code that will do these questions is as follows.

```
#
#  This is code to use the EDUC data,
#  test for heteroskedasticity and correct for it
#
```

```
rm(list=ls(all=TRUE))                # remove all objects in memory

#  READ DATA
educ = read.csv("educ.csv", head=TRUE, sep=",")
attach(educ)
N = length(EDUC)                     # sample length

#  OLS REGRESSION
reg = lm(EDUC ~ GDP + POP)
print(summary(reg))

#  QUESTION 2a
# GOLDFELD-QUANDT TEST USING GDP OR POP
g=order(GDP)                         # PUT EITHER GDP OR POP IN THE BRACKETS
EDUCg = EDUC[g]                      #
GDPg = GDP[g]                        # SORT THE DATA BY g
POPg = POP[g]                        #
                        # RUN REGS LEAVING OUT MIDDLE 8
reg_lo = lm(EDUCg[1:15]  ~ GDPg[1:15]  + POPg[1:15])
reg_hi = lm(EDUCg[24:38] ~ GDPg[24:38] + POPg[24:38])
                        # GQ STAT
GQ_stat = sum(reg_hi$residuals^2)/sum(reg_lo$residuals^2)
p_val = 1-pf(GQ_stat, 12, 12)
cat("The Goldfeld-Quandt statistic is", GQ_stat, "which has a p value of",
   p_val, "\n")


#  QUESTION 2b
#  BREUSCH-PAGAN TEST
sighat_sq = sum(reg$residuals^2)/N  # FORM DEPENDENT VARIABLE
eps_hat = reg$residuals^2/sighat_sq #
BPreg = lm(eps_hat ~ GDP + POP)      # BP REGRESSION
RSS = sum( (BPreg$fitted - mean(eps_hat))^2 )
BP_stat = RSS/2                      # BP STAT
p_val = 1-pchisq(BP_stat, 2)
cat("The Breusch-Pagan statistic is", BP_stat, "which has a p value of",
   p_val, "\n")

#  QUESTION 2c
#  WHITE TEST
epsq = reg$residuals^2               # FORM DEPENDENT VARIABLE
GDP2 = GDP^2                         #
POP2 = POP^2                         # SQUARES AND
GDPPOP = GDP*POP                     # CROSS-PRODUCT TERM
Wreg = lm(epsq~GDP+POP+GDP2+POP2+GDPPOP)  # WHITE REGRESSION
W_stat = N*summary(Wreg)$r.squared
p_val = 1-pchisq(W_stat, 2)          # WHITE STAT
cat("The Whites statistic is", W_stat, "which has a p value of", p_val, "\n")

#  QUESTION 4
#  GLS MODEL
invGDP = 1/GDP
POP_GDP = POP/GDP
EDUC_GDP = EDUC/GDP
greg = lm(EDUC_GDP ~ invGDP+POP_GDP)
print(summary(greg))
```

```
#   QUESTION 5
#   HAC ESTIMATION
library(sandwich)
vv = vcovHC(reg)                    # COMPUTES WHITE's V-COV MATRIX
   sqrt_vv = chol(vv)               # TAKES SQUARE ROOT VCOV MATRIX
   sd = diag(sqrt_vv)               # TAKES DIAGONAL ELEMENTS
   new_t = reg$coef / sd            # FORMS NEW T STATS
   cat("Het. and Autocorrelation-Consistent t-stats:", "\n")
   print(new_t)
```

2 (e) In this case 2 of the 3 test scores reject homoskedasticity, while the 3rd (Breusch-Pagan) only weakly rejects it (p>0.05). So if you only did the BP test you might not detect the heteroskedasticity.

3. You'll get a test score of 3.41 (p = 0.18). So you would not detect the heteroskedasticity in this case.

4. F, F, T, T, T

Unit 10

1. We will use the fact that if $0 < c < 1$ then $\sum_{t=0}^{\infty} c^t = 1/(1-c)$.

$$V(e_t) = V(u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + \cdots)$$

$$= V\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right)$$

$$= \sum_{i=0}^{\infty} \rho^{2i} V(u_{t-i})$$

$$= \frac{\sigma^2}{1-\rho^2}.$$

2. Note that $u_t = e_t - \rho e_{t-1}$. Multiply the regression model through by ρ and lag one period to obtain $\rho y_{t-1} = \alpha\rho + \beta\rho X_{t-1} + \rho e_{t-1}$. Subtract that from the original regression to get

$$y_t - \rho y_{t-1} = \alpha - \alpha\rho + \beta(X_t - \rho X_{t-1}) + e_t - \rho e_{t-1}.$$

Then our regression is $y_t^* = \alpha^* + \beta X_t^* + u_t$ where $y_t^* = y_t - \rho y_{t-1}$, $\alpha^* = \alpha - \alpha\rho$, X_t^*=$(X_t - \rho X_{t-1})$ and $V(u_t) = \sigma^2$.

3. The slope coefficient is 0.954 and it is significant (p = ~0).

4. The results will be

```
LM Test score for AR1 = 67.10104 with p value 2.220446e-16
LM Test score for AR2 = 68.83038 with p value 1.110223e-15
```

Since both scores are significant there is evidence of at least AR2 errors. Since the p values are so low, there is likely even higher order autocorrelation.

5. The slope coefficient drops from 0.954 to 0.542 but remains significant.

6. The Newey-West standard errors are (constant) 0.6323 and (sales) 0.1306.
   The HAC t-statistics are 0.4504 and 7.7066.

Code to compute all these answers is as follows.

```
#
#  This is code to use the COMPUTER data for Unit 10,
#  test for serial correlation and correct for it
#

rm(list=ls(all=TRUE))                # remove all objects in memory


#  QUESTION 3

# READ DATA
cdata = read.csv("computer.csv", head=FALSE, skip=1, sep=",")
attach(cdata)
T = length(V1)                       # sample length
sales = V2                           # % growth in sales
comp = V3                            # % change in computer purchases
# OLS REGRESSION
ols = lm(sales ~ comp)
print(summary(ols))


#  QUESTION 4
# LM Tests for autocorrelation
lag_e1 = c(0,ols$residuals[1:(T-1)])      # generate lagged residuals
lag_e2= c(0,lag_e1[1:(T-1)])
# AR1 TEST
lm1 = lm(ols$residual[2:T] ~ comp[2:T] + lag_e1[2:T] )
   lmtest = T*summary(lm1)$r.squared      # Test score & pvalue
   pvalue = 1-pchisq(lmtest, 1)
   cat("LM Test score for AR1 =", lmtest, "with p value", pvalue, "\n")
# AR2 TEST
lm2 = lm(ols$residual[3:T] ~ comp[3:T] + lag_e1[3:T] + lag_e2[3:T])
   summary(lm2)
   lmtest = T*summary(lm2)$r.squared      # Test score & pvalue
   pvalue = 1-pchisq(lmtest, 2)
   cat("LM Test score for AR2 =", lmtest, "with p value", pvalue, "\n")


#  QUESTION 5
# Cochrane-Orcutt estimation
library(orcutt)
co = cochrane.orcutt(ols)
print(summary(co))
cat("\n","The rho coefficient is", co$rho,"\n")
cat("\n","\n")


#  QUESTION 6
#  HAC ESTIMATION
library(sandwich)
```

```
vv = NeweyWest(ols)                    # COMPUTES NW V-COV MATRIX
   sqrt_vv = chol(vv)                  # TAKES SQUARE ROOT VCOV MATRIX
   sd = diag(sqrt_vv)                  # TAKES DIAGONAL ELEMENTS
   new_t = ols$coef / sd                     # FORMS NEW T STATS
   cat("Newey-West std errors:", "\n")
   print(sd)
   cat("\n", "Het. and Autocorrelation-Consistent t-stats:", "\n")
   print(new_t)
```

7. F, T, F, T, T

Unit 11

1. Rearrange the second equation to get $-\gamma Y_i = -X_i + e_{2i}$ which implies

$$Y_i = \frac{1}{\gamma}X_i - \frac{1}{\gamma}e_{2i}.$$

Now set the two equations for Y_i equal to each other and solve for X_i to get

$$X_i = \frac{e_{1i} + \frac{1}{\gamma}e_{2i}}{\frac{1}{\gamma} - \beta}.$$

Take the covariance: $cov(X_i e_{1i}) = E(X_i e_{1i}) = E\left(\frac{e_{1i}+\frac{1}{\gamma}e_{2i}}{\frac{1}{\gamma}-\beta}e_{1i}\right) = \frac{1}{\frac{1}{\gamma}-\beta}E\left(e_{1i}^2\right) = \frac{\gamma\sigma_1^2}{1-\gamma\beta} \neq 0.$

Note this result uses the fact that the means of both X and e are zero.

2. To compute the probability limit of $\bar{X}$ we need to find the expected value and the variance. Using standard rules we have $E(aX) = a\mu$, $var(aX) = a^2\sigma^2$, $E(a\bar{X}) = a\mu$ and $var(a\bar{X}) = \frac{a^2}{N}\sigma^2$. (See Unit 1 to remind yourself of these rules). The sample size N doesn't appear in the expression $E(a\bar{X})$ so as $N \to \infty$ it remains $a\mu$. But N appears in the denominator of the variance so as $N \to \infty$, $var(a\bar{X}) \to 0$. Thus $plim(a\bar{X}) = a\mu$.

3. The results using just SM as an instrument are

```
Call:
ivreg(formula = LogEARN ~ TEST + S + MALE + ETHBLACK + ETHHISP |
    S + MALE + ETHBLACK + ETHHISP + SM)

Residuals:
      Min        1Q    Median        3Q       Max
-2.397411 -0.302299  0.003193  0.312006  1.518342

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.33428    0.62655   0.534    0.594
TEST         0.03649    0.02286   1.596    0.111
S            0.03045    0.04329   0.703    0.482
MALE         0.26467    0.05783   4.577 5.87e-06 ***
ETHBLACK     0.17289    0.25493   0.678    0.498
ETHHISP      0.16980    0.18324   0.927    0.355

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   1 534    11.300 0.000831 ***
Wu-Hausman         1 533     1.373 0.241867
Sargan             0  NA        NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5341 on 534 degrees of freedom
Multiple R-Squared: 0.1991,     Adjusted R-squared: 0.1916
Wald test:  36.2 on 5 and 534 DF,  p-value: < 2.2e-16
```

Adding SF and SIBLINGS as instruments gives

```
Call:
ivreg(formula = LogEARN ~ TEST + S + MALE + ETHBLACK + ETHHISP |
    S + MALE + ETHBLACK + ETHHISP + SM + SF + SIBLINGS)

Residuals:
      Min        1Q    Median        3Q       Max
-2.398915 -0.307556  0.002477  0.326477  1.548063

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.23415    0.53046   0.441   0.6591
TEST         0.04022    0.01917   2.099   0.0363 *
S            0.02353    0.03667   0.642   0.5213
MALE         0.25895    0.05525   4.687 3.52e-06 ***
ETHBLACK     0.21189    0.21981   0.964   0.3355
ETHHISP      0.19538    0.16310   1.198   0.2315

Diagnostic tests:
                 df1 df2 statistic  p-value
Weak instruments   3 532     5.579 0.000897 ***
Wu-Hausman         1 533     2.702 0.100817
Sargan             2  NA     0.920 0.631286
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5432 on 534 degrees of freedom
Multiple R-Squared: 0.1714,     Adjusted R-squared: 0.1636
Wald test: 35.37 on 5 and 534 DF,  p-value: < 2.2e-16
```

The coefficient on TEST gets slightly larger and becomes significant. However the Wu-Hausman test still fails to reject, although the *p* value gets smaller.

4. The fitted values will be $\hat{\gamma} Z_i$. Call the new estimator $\tilde{\beta}_{IV}$

$$\tilde{\beta}_{IV} = \frac{\Sigma\hat{\gamma}Z_iY_i}{\Sigma X_i\hat{\gamma}Z_i} = \frac{\hat{\gamma}\Sigma Z_iY_i}{\hat{\gamma}\Sigma X_iZ_i} = \frac{\Sigma Z_iY_i}{\Sigma X_iZ_i} = \hat{\beta}_{IV}$$

The two estimators will yield the same result.

5. T, T, T, F, T

Unit 12

1. The dot product is $4 \times -2 + 2 \times 7 - 3 \times 2 = 0$.

2. $\begin{bmatrix} 6 & 66 \\ -37 & 18 \end{bmatrix}$

3. $\begin{bmatrix} 9 & -3 \\ -1 & 6 \end{bmatrix}\begin{bmatrix} 2 & -4 \\ 12 & 1 \end{bmatrix} = \begin{bmatrix} -18 & -39 \\ 70 & 10 \end{bmatrix}$

4. 100x12; 13x13; 13x100.

5. $Y = Xb + e$ where Y and e are each N-length vectors, $X = \begin{bmatrix} 1 & x_{11} \cdots & x_{14} \\ \vdots & \vdots & \vdots \\ 1 & x_{N1} \cdots & x_{N4} \end{bmatrix}$ is an Nx5 matrix and $b^T = [b_0\ b_1\ b_2\ b_3\ b_4]$ is a parameter vector.

6. $(k \times N)(N \times k) \times (k \times N)(N \times 1) \rightarrow (k \times k)(k \times 1) \rightarrow k \times 1$.

7. They are the standard errors of the regression coefficients.

```
> sqrt(diag(vcov))
(Intercept)          X1          X2          X3
 10.6005323   0.2481388   0.2166730   1.1065259
```