Name Mr. Key
ID 7777777

ECO 5350
Intro. Econometrics

Prof. T. Fomby
Spring 2011

Mid-Term Again
Exam

Instructions: Put your name and student ID in the upper right-hand-corner of this exam. This exam is worth a total of 50 points. The breakout of these points by question is as follows:

Q1 = (3,3,3,2) = 11 points
Q2 = (3,5) = 8 points
Q3 = (10,4) = 14 points
Q4 = (5,2,4,2,4) = 17 points

You have one hour and thirty minutes to take this test. Don't get hung up on any one question. Answer the easy questions first and then go back and pick up the hard ones. Good luck.

1. This question covers concepts concerning the **Security Characteristic Line** and **Jensen's alpha** in financial economics. In this case y = gm– rkfree, x = mkt – rkfree where gm stands for monthly returns to General Motors stock, rkfree is the monthly rate of return on one-month t-bills, and mkt is the monthly rate of return on the S&P 500 stock index. Use the following EVIEWS output to answer the following questions:

Dependent Variable: Y
Method: Least Squares
Date: 03/29/11 Time: 21:23
Sample: 1998M01 2008M12
Included observations: 132

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.007248	0.011393	-0.636145	0.5258
X	1.146838	0.197242	5.814356	0.0000

R-squared	0.206382	Mean dependent var	-0.041185
Adjusted R-squared	0.200277	S.D. dependent var	0.125706
S.E. of regression	0.112415	Akaike info criterion	-1.518201
Sum squared resid	1.642830	Schwarz criterion	-1.474522
Log likelihood	102.2013	Hannan-Quinn criter.	-1.500452
F-statistic	33.80673	Durbin-Watson stat	2.035563
Prob(F-statistic)	0.000000		

3

a. Jensen's alpha $\hat{\alpha}$ in this case is <u>−0.007248</u>. Its standard error is <u>0.011393</u>. Its t-statistic is <u>−0.636145</u>

3

b. Compute a **90%** confidence interval for Jensen's alpha. It is [<u>−0.02598</u>, <u>0.01149</u>]. Show your work below.

Formula: $b_1 \pm se(b_1) t_c$

$t_c = t_{(0.95, \infty)} = 1.645$

$-0.007248 \pm 0.011393 (1.645)$

[−0.0 [redacted] 2598, 0.01149]

c. (Students: Be careful with this question.) Jane Success is a contrarian investor who looks for stocks that have **significantly inferior** risk-adjusted rates of return. The reason she does so is because she believes that the stock market is not rational and tends, for substantial periods of time, to run down the price of "out-of-favor" stocks. Using a **10% level of significance,** test the appropriate null and alternative hypotheses that would allow Jane to determine if she should invest in the GM stock. Should she invest in GM stock? Clearly explain your answer.

If one wants to pursue

we want to test $H_0: \beta_1 \geq 0$ versus $H_1: \beta_1 < 0$.

Acceptance Region = $(-1.282, \infty)$

Rejection Region = $(-\infty, -1.282)$

3

The t-statistic for this test is −0.636145 which is already given in the above computer output. This is a one-tailed test therefore the one-tailed p-value is ~~0.5258~~ > 0.10. Therefore, we accept H_0. GM is not statistically inferior and Jane should not invest in it.

one-tailed p-value = 0.2629 > 0.10

d. Is ~~Microsoft~~ GM stock a conservative stock or an aggressive stock? Explain your answer.

(2) It is an aggressive stock as it "beta" (= 1.146838) is greater than 1.0. It price is more volatile than the overall stock market as represented by the S&P 500 index.

2. The file *toodyay.dat* contains 48 annual observations on a number of variables related to wheat yield in the Toodyay Shire of Western Australia, for the period 1950 – 1997. Those variables are

Y = wheat yield in tons per hectare,
t = trend term to allow for technological change,
RG = rainfall at **germination** (May – June)
RD = rainfall at **development** stage (July – August)
RF = rainfall at **flowering** (September – October)

The unit of measurement for rainfall is centimeters. A model that allows for the yield response to rainfall to be different for the three different periods is

$$Y = \beta_1 + \beta_2 t + \beta_3 RG + \beta_4 RD + \beta_5 RF + e\,.$$

Consider following two least squares regressions involving this data.

Dependent Variable: Y
Method: Least Squares
Date: 03/29/11 Time: 19:29
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.625380	0.258218	2.421908	0.0197
T	0.030210	0.003439	8.785445	0.0000
RG	-0.079423	0.081693	-0.972207	0.3364
RD	-0.000480	0.091763	-0.005230	0.9959
RF	0.338711	0.165429	2.047474	0.0468

R-squared	0.688875	Mean dependent var	1.476737
Adjusted R-squared	0.659933	S.D. dependent var	0.542493
S.E. of regression	0.316357	Akaike info criterion	0.634439
Sum squared resid	4.303504	Schwarz criterion	0.829356
Log likelihood	-10.22653	Hannan-Quinn criter.	0.708098
F-statistic	23.80199	Durbin-Watson stat	1.094043
Prob(F-statistic)	0.000000		

Two tailed rejection region → $(-\infty, -2.021) \cup (2.021, \infty)$

Dependent Variable: Y
Method: Least Squares
Date: 03/29/11 Time: 19:37
Sample: 1 48
Included observations: 48

part a.
I accepted both one tailed and two-tailed answers as long as you did them correctly.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.464614	0.145388	3.195677	0.0026
T	0.031068	0.003258	9.536747	0.0000
RF	0.349625	0.159110	2.197381	0.0332

R-squared	0.682029	Mean dependent var	1.476737
Adjusted R-squared	0.667897	S.D. dependent var	0.542493
S.E. of regression	0.312630	Akaike info criterion	0.572869
Sum squared resid	4.398188	Schwarz criterion	0.689819
Log likelihood	-10.74885	Hannan-Quinn criter.	0.617064
F-statistic	48.26125	Durbin-Watson stat	0.989452
Prob(F-statistic)	0.000000		

If you consider rejection and acceptance regions we have accept $(-\infty, 1.684)$ reject $(1.684, \infty)$

a. Consider the **first** regression output above. Does it appear that there has been a significant technical change in the production of wheat in Australia? **Use a 5% level of significance**. Thoroughly explain your answer. What we need to do is test the statistical significance of the time trend variable "t".

3

$H_0: \beta_2 = 0$ versus $H_1: \beta_2 > 0$. This is a one-sided test. If $\beta_2 < 0$ then we have had a technical decline in the country. The t-statistic of interest is 8.785 whose one-sided p-value is $0.0000/2 = 0.0000 < 0.05$ and therefore, the has been positive technical change over time.

b. According to veteran wheat farmers in Australia, the amount of rainfall that occurs during the **germination** and **development** stages is not nearly as important as the rainfall that occurs during the **flowering** stage. Test this hypothesis using a subset F-test. Clearly state the null hypothesis that supports the Australian wheat farmers' claim and the corresponding alternative hypothesis that refutes their claim. Test the proposition at the **5% level of significance**. Here we are the "subset" F-test

$H_0: \beta_3 = \beta_4 = 0$ versus $H_1: \beta_3 \neq 0, \beta_4 \neq 0$, or both. The F-statistic

5

is calculated as $F = \dfrac{(SSE_r - SSE_u)/J}{SSE_u/(N-K)} = \dfrac{(4.398188 - 4.303504)/2}{4.303504/(48-5)}$

$= \dfrac{0.047342}{4.303504/43} = \dfrac{0.047342}{0.10008} = 0.473$. The critical region for the test is approximately $(3.20, \infty)$. The acceptance region is therefore $(0, 3.20)$. Since the observed F-statistic falls in the acceptance region, we accept the proposition that rainfall during the germination and development stages are not statistically important.

3. a. Use the **first regression output of question 2** to fill in the following ANOVA table.

Source	SS	DF	MS	F(overall)
Model	9.5285	4 $K-1$	2.382	23.80 (p= 0.00)
Error	4.303504	43 $N-K$	0.10008	
Total	13.8320	47 $N-1$		

10

Show your work to receive full credit. Straight from the output we have $SSE = 4.303504$, $R^2 = 0.688875$, $F_{overall} = 23.80$ with p-value $= 0.0000$. Recall $R^2 = 1 - \frac{SSE}{TSS}$, therefore, $1 - R^2 = SSE/TSS$ and $TSS = \frac{SSE}{1-R^2} = \frac{4.303504}{1-0.688875} = 13.8320$

$SSM = TSS - SSE = 13.8320 - 4.303504 = 9.5285$

b. What is the null hypothesis that the above table is testing? It is testing the joint significance of all of the explanatory variables in the model excluding the intercept i.e. $H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ versus H_1: not H_0. Here the F-statistic is highly significant as its p-value is much less than 0.05. Thus the variables T, RG, RD, and RF are jointly statistically significant.

4

4. To examine the quantity theory of money, Brumm in a Southern Economics Journal article in 2005 proposed the following regression equation

$$INFLAT = \beta_1 + \beta_2 MONEY + \beta_3 OUTPUT + e$$

where INFLAT is the growth rate in the general price level, MONEY is the growth rate of the money supply, and OUTPUT is the growth rate in the national output. All growth rates are measured in terms of percentage.

a. One of the so-called weak hypotheses of the quantity theory of money is that

$$H_0: \beta_2 + \beta_3 = 0 \text{ versus } H_1: \beta_2 + \beta_3 \neq 0$$

Use the following two computer outputs to test the above hypotheses at the **5% level of significance**. Answer your question in the space below the computer output.

Dependent Variable: INFLAT
Method: Least Squares
Date: 03/29/11 Time: 23:23
Sample: 1 76
Included observations: 76

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.234214	0.979925	-0.239012	0.8118
MONEY	1.033131	0.009042	114.2565	0.0000
OUTPUT	-1.662006	0.250566	-6.633003	0.0000

R-squared	0.994797	Mean dependent var	25.35395
Adjusted R-squared	0.994654	S.D. dependent var	58.94767
S.E. of regression	4.309966	Akaike info criterion	5.798411
Sum squared resid	1356.034	Schwarz criterion	5.890413
Log likelihood	-217.3396	Hannan-Quinn criter.	5.835179
F-statistic	6978.325	Durbin-Watson stat	2.305899
Prob(F-statistic)	0.000000		

Covariance Matrix of least squares estimates:

	C	MONEY	OUTPUT
C	0.960253	-0.003774	-0.201562
MONEY	-0.003774	8.18E-05	0.000452
OUTPUT	-0.201562	0.000452	0.062783

The t-statistic we are interested in is

5

$$t = \frac{b_2 + b_3 - 0}{se(b_2+b_3)} = \frac{1.033131 - 1.662006}{0.2525} = \frac{-0.628875}{0.2525} = -2.4905$$

Here $se(b_2+b_3) = \sqrt{\hat{var}(b_2) + \hat{var}(b_3) + 2\hat{cov}(b_2, b_3)}$

$$= \sqrt{0.0000818 + 0.062783 + 2(0.000452)} = 0.2525$$

Here we have a two-sided test. $H_0: \beta_2 + \beta_3 = 0$ versus $H_1: \beta_2 + \beta_3 \neq 0$. The acceptance region for this test is $(-t_c, t_c)$ while the rejection region is $(-\infty, -t_c) \cup (t_c, \infty)$ where $t_c = t_{0.975, \infty} = 1.96$. Since the observed value of the t-statistic, -2.4905, falls in the critical region, we reject the null hypothesis of the "weak" form of the quantity theory of money.

Consider the following "shifted" regression equation where MONEY_S = MONEY – 8 and OUTPUT_S = OUTPUT – 2.

Dependent Variable: INFLAT
Method: Least Squares
Date: 03/29/11 Time: 23:55
Sample: 1 76
Included observations: 76

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.706823	0.603697	7.796664	0.0000
MONEY_S	1.033131	0.009042	114.2565	0.0000
OUTPUT_S	-1.662006	0.250566	-6.633003	0.0000

R-squared	0.994797	Mean dependent var	25.35395
Adjusted R-squared	0.994654	S.D. dependent var	58.94767
S.E. of regression	4.309966	Akaike info criterion	5.798411
Sum squared resid	1356.034	Schwarz criterion	5.890413
Log likelihood	-217.3396	Hannan-Quinn criter.	5.835179
F-statistic	6978.325	Durbin-Watson stat	2.305899
Prob(F-statistic)	0.000000		

(2) b. Suppose that we are interested in predicting the **expected** inflation rate, given the money supply has grown at 8% while the output in the national economy has only grown at 2%. Your BLU prediction of the **expected** inflation rate is 4.70690.

This is given directly from the "translated" output above.

(4) c. Construct a 95% confidence interval for the **expected** inflation rate given MONEY = 8% and GROWTH (OUTPUT) = 2%. Show you work if you expect full credit. The 95% confidence interval is [3.522, 5.889]. The desired C.I. is $4.706 \pm se(\cdot)t_c$ where $se(\cdot)$ is given as the standard error of the intercept in the above translated regression. Therefore $4.706 \pm 0.603697(1.96)$

$= [3.522, 5.889]$

(2) d. Suppose that we are interested in predicting the inflation rate **for the next period**, given the money supply has grown at 8% while the output in the national economy has only grown at 2%. Your BLU prediction of the inflation rate **for the next period** is 4.706. The BLU predictions for the mean of the inflation rates and next inflation rate are the same.

(4) e. Construct a 95% confidence interval for the inflation rate **for the next period**, given MONEY = 8% and GROWTH (OUTPUT) = 2%. Show you work if you expect full credit. The 95% confidence interval is [-3.823, 13.235]. Again the desired C.I. is of the form $4.706 \pm se(\cdot)t_c$. This time, however, $se(\hat{y}_0) =$

$\sqrt{0.603697^2 + 4.309966^2} = 4.352$. Therefore, the desired C.I. is $4.706 \pm (4.352)(1.96) = [-3.823, 13.235]$

The Rules of Summation

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

$$\sum_{i=1}^{n} a = na$$

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} (ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} (a + bx_i) = na + b\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

$$\begin{aligned}\sum_{i=1}^{2}\sum_{j=1}^{3} f(x_i, y_j) &= \sum_{i=1}^{2} [f(x_i, y_1) + f(x_i, y_2) + f(x_i, y_3)] \\ &= f(x_1, y_1) + f(x_1, y_2) + f(x_1, y_3) \\ &\quad + f(x_2, y_1) + f(x_2, y_2) + f(x_2, y_3)\end{aligned}$$

Expected Values & Variances

$$\begin{aligned}E(X) &= x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) \\ &= \sum_{i=1}^{n} x_i f(x_i) = \sum_x x f(x)\end{aligned}$$

$$E[g(X)] = \sum_x g(x) f(x)$$

$$\begin{aligned}E[g_1(X) + g_2(X)] &= \sum_x [g_1(x) + g_2(x)] f(x) \\ &= \sum_x g_1(x) f(x) + \sum_x g_2(x) f(x) \\ &= E[g_1(X)] + E[g_2(X)]\end{aligned}$$

$E(c) = c$
$E(cX) = cE(X)$
$E(a + cX) = a + cE(X)$
$\text{var}(X) = \sigma^2 = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$
$\text{var}(a + cX) = E[(a + cX) - E(a + cX)]^2 = c^2\text{var}(X)$

Marginal and Conditional Distributions

$f(x) = \sum_y f(x, y)$ for each value X can take

$f(y) = \sum_x f(x, y)$ for each value Y can take

$$f(x|y) = P[X = x | Y = y] = \frac{f(x, y)}{f(y)}$$

If X and Y are independent random variables, then $f(x,y) = f(x)f(y)$ for each and every pair of values x and y. The converse is also true.

If X and Y are independent random variables, then the conditional probability density function of X given that

$$Y = y \text{ is } f(x|y) = \frac{f(x, y)}{f(y)} = \frac{f(x)f(y)}{f(y)} = f(x)$$

for each and every pair of values x and y. The converse is also true.

Expectations, Variances & Covariances

$$\begin{aligned}\text{cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= \sum_x\sum_y [x - E(X)][y - E(Y)] f(x, y)\end{aligned}$$

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

$E(c_1X + c_2Y) = c_1E(X) + c_2E(Y)$
$E(X + Y) = E(X) + E(Y)$

$\text{var}(aX + bY + cZ) = a^2\text{var}(X) + b^2\text{var}(Y) + c^2\text{var}(Z) + 2ab\text{cov}(X,Y) + 2ac\text{cov}(X,Z) + 2bc\text{cov}(Y,Z)$

If X, Y, and Z are independent, or uncorrelated, random variables, then the covariance terms are zero and:

$$\text{var}(aX + bY + cZ) = a^2\text{var}(X) + b^2\text{var}(Y) + c^2\text{var}(Z)$$

Normal Probabilities

If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$

If $X \sim N(\mu, \sigma^2)$ and a is a constant, then

$$P(X \geq a) = P\left(Z \geq \frac{a - \mu}{\sigma}\right)$$

If $X \sim N(\mu, \sigma^2)$ and a and b are constants, then

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

Assumptions of the Simple Linear Regression Model

- SR1 The value of y, for each value of x, is $y = \beta_1 + \beta_2 x + e$
- SR2 The average value of the random error e is $E(e) = 0$ since we assume that $E(y) = \beta_1 + \beta_2 x$
- SR3 The variance of the random error e is $\text{var}(e) = \sigma^2 = \text{var}(y)$
- SR4 The covariance between any pair of random errors, e_i and e_j is $\text{cov}(e_i, e_j) = \text{cov}(y_i, y_j) = 0$
- SR5 The variable x is not random and must take at least two different values.
- SR6 (*optional*) The values of e are *normally distributed* about their mean $e \sim N(0, \sigma^2)$

Least Squares Estimation

If b_1 and b_2 are the least squares estimates, then

$$\hat{y}_i = b_1 + b_2 x_i$$
$$\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i$$

The Normal Equations

$$Nb_1 + \Sigma x_i b_2 = \Sigma y_i$$
$$\Sigma x_i b_1 + \Sigma x_i^2 b_2 = \Sigma x_i y_i$$

Least Squares Estimators

$$b_2 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$b_1 = \bar{y} - b_2\bar{x}$$

Elasticity

$$\eta = \frac{\text{percentage change in } y}{\text{percentage change in } x} = \frac{\Delta y/y}{\Delta x/x} = \frac{\Delta y}{\Delta x}\cdot\frac{x}{y}$$

$$\eta = \frac{\Delta E(y)/E(y)}{\Delta x/x} = \frac{\Delta E(y)}{\Delta x}\cdot\frac{x}{E(y)} = \beta_2\cdot\frac{x}{E(y)}$$

Least Squares Expressions Useful for Theory

$$b_2 = \beta_2 + \Sigma w_i e_i$$

$$w_i = \frac{x_i - \bar{x}}{\Sigma(x_i - \bar{x})^2}$$

$$\Sigma w_i = 0,\quad \Sigma w_i x_i = 1,\quad \Sigma w_i^2 = 1/\Sigma(x_i - \bar{x})^2$$

Properties of the Least Squares Estimators

$$\text{var}(b_1) = \sigma^2\left[\frac{\Sigma x_i^2}{N\Sigma(x_i - \bar{x})^2}\right] \quad \text{var}(b_2) = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}$$

$$\text{cov}(b_1, b_2) = \sigma^2\left[\frac{-\bar{x}}{\Sigma(x_i - \bar{x})^2}\right]$$

Gauss-Markov Theorem: Under the assumptions SR1–SR5 of the linear regression model the estimators b_1 and b_2 have the *smallest variance of all linear and unbiased estimators* of β_1 and β_2. They are the Best Linear Unbiased Estimators (BLUE) of β_1 and β_2.

If we make the normality assumption, assumption SR6, about the error term, then the least squares estimators are normally distributed.

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2\Sigma x_i^2}{N\Sigma(x_i - \bar{x})^2}\right), b_2 \sim N\left(\beta_2, \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}\right)$$

Estimated Error Variance

$$\hat{\sigma}^2 = \frac{\Sigma\hat{e}_i^2}{N-2}$$

Estimator Standard Errors

$$\text{se}(b_1) = \sqrt{\widehat{\text{var}(b_1)}},\quad \text{se}(b_2) = \sqrt{\widehat{\text{var}(b_2)}}$$

***t*-distribution**

If assumptions SR1–SR6 of the simple linear regression model hold, then

$$t = \frac{b_k - \beta_k}{\text{se}(b_k)} \sim t_{(N-2)},\quad k = 1,2$$

Interval Estimates

$$P[b_2 - t_c\text{se}(b_2) \le \beta_2 \le b_2 + t_c\text{se}(b_2)] = 1 - \alpha$$

Hypothesis Testing

Components of Hypothesis Tests

1. A *null* hypothesis, H_0
2. An *alternative* hypothesis, H_1
3. A test *statistic*
4. A *rejection* region
5. A conclusion

If the null hypothesis $H_0: \beta_2 = c$ is ***true***, then

$$t = \frac{b_2 - c}{\text{se}(b_2)} \sim t_{(N-2)}$$

Rejection rule for a two-tail test: If the value of the test statistic falls in the rejection region, either tail of the t-distribution, then we reject the null hypothesis and accept the alternative.

Type I error: The null hypothesis is *true* and we decide to *reject* it.

Type II error: The null hypothesis is *false* and we decide *not* to reject it.

***p*-value rejection rule:** When the p-value of a hypothesis test is *smaller* than the chosen value of α, then the test procedure leads to *rejection* of the null hypothesis.

Prediction

$$y_0 = \beta_1 + \beta_2 x_0 + e_0,\ \hat{y}_0 = b_1 + b_2 x_0,\ f = \hat{y}_0 - y_0$$

$$\widehat{\text{var}(f)} = \hat{\sigma}^2\left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right], \text{se}(f) = \sqrt{\widehat{\text{var}(f)}}$$

A $(1-\alpha)\times 100\%$ confidence interval, or prediction interval, for y_0

$$\hat{y}_0 \pm t_c\text{se}(f)$$

Goodness of Fit

$$\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma\hat{e}_i^2$$

$$SST = SSR + SSE$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = (\text{corr}(y, \hat{y}))^2$$

Log-Linear Model

$$\ln(y) = \beta_1 + \beta_2 x + e,\ \widehat{\ln(y)} = b_1 + b_2 x$$

$100 \times \beta_2 \approx$ % change in y given a one-unit change in x.

$$\hat{y}_n = \exp(b_1 + b_2 x)$$

$$\hat{y}_c = \exp(b_1 + b_2 x)\exp(\hat{\sigma}^2/2)$$

Prediction interval:

$$\exp\left[\widehat{\ln(y)} - t_c\text{se}(f)\right],\ \exp\left[\widehat{\ln(y)} + t_c\text{se}(f)\right]$$

Generalized goodness-of-fit measure $R_g^2 = (\text{corr}(y, \hat{y}_n))^2$

Assumptions of the Multiple Regression Model

MR1 $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i$

MR2 $E(y_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} \Leftrightarrow E(e_i) = 0.$

MR3 $\text{var}(y_i) = \text{var}(e_i) = \sigma^2$

MR4 $\text{cov}(y_i, y_j) = \text{cov}(e_i, e_j) = 0$

MR5 The values of x_{ik} are not random and are not exact linear functions of the other explanatory variables.

MR6 $y_i \sim N[(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}), \sigma^2] \Leftrightarrow e_i \sim N(0, \sigma^2)$

Least Squares Estimates in MR Model

Least squares estimates $b_1, b_2, \ldots, b_K$ minimize
$S(\beta_1, \beta_2, \ldots, \beta_K) = \Sigma(y_i - \beta_1 - \beta_2 x_{i2} - \cdots - \beta_K x_{iK})^2$

Estimated Error Variance and Estimator Standard Errors

$$\hat{\sigma}^2 = \frac{\Sigma\hat{e}_i^2}{N-K} \quad \text{se}(b_k) = \sqrt{\widehat{\text{var}(b_k)}}$$

Hypothesis Tests and Interval Estimates for Single Parameters

Use t-distribution $t = \dfrac{b_k - \beta_k}{\mathrm{se}(b_k)} \sim t_{(N-K)}$

t-test for More than One Parameter

$$H_0: \beta_2 + c\beta_3 = a$$

When H_0 is true $t = \dfrac{b_2 + cb_3 - a}{\mathrm{se}(b_2 + cb_3)} \sim t_{(N-K)}$

$$\mathrm{se}(b_2 + cb_3) = \sqrt{\widehat{\mathrm{var}(b_2)} + c^2\widehat{\mathrm{var}(b_3)} + 2c \times \widehat{\mathrm{cov}(b_2, b_3)}}$$

Joint F-tests

To test J joint hypotheses,

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N-K)}$$

To test the overall significance of the model the null and alternative hypotheses and F statistic are

$$H_0: \beta_2 = 0,\ \beta_3 = 0, \ldots, \beta_K = 0$$
H_1 : *at least one* of the β_k is nonzero

$$F = \frac{(SST - SSE)/(K-1)}{SSE/(N-K)}$$

RESET: A Specification Test

$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \qquad \hat{y}_i = b_1 + b_2 x_{i2} + b_3 x_{i3}$

$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma_1 \hat{y}_i^2 + e_i, \quad H_0: \gamma_1 = 0$

$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + e_i, \quad H_0: \gamma_1 = \gamma_2 = 0$

Model Selection

$\mathrm{AIC} = \ln(SSE/N) + 2K/N$

$\mathrm{SC} = \ln(SSE/N) + K\ln(N)/N$

Collinearity and Omitted Variables

$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$

$$\mathrm{var}(b_2) = \frac{\sigma^2}{(1 - r_{23}^2)\sum(x_{i2} - \bar{x}_2)^2}$$

When x_3 is omitted, $\mathrm{bias}(b_2^*) = E(b_2^*) - \beta_2 = \beta_3 \dfrac{\widehat{\mathrm{cov}(x_2, x_3)}}{\widehat{\mathrm{var}(x_2)}}$

Heteroskedasticity

$\mathrm{var}(y_i) = \mathrm{var}(e_i) = \sigma_i^2$

General variance function

$\sigma_i^2 = \exp(\alpha_1 + \alpha_2 z_{i2} + \cdots + \alpha_S z_{iS})$

Breusch-Pagan and White Tests for H_0: $\alpha_2 = \alpha_3 = \cdots = \alpha_S = 0$

When H_0 is true $\chi^2 = N \times R^2 \sim \chi^2_{(S-1)}$

Goldfeld-Quandt test for $H_0: \sigma_M^2 = \sigma_R^2$ versus $H_1: \sigma_M^2 \neq \sigma_R^2$

When H_0 is true $F = \hat{\sigma}_M^2/\hat{\sigma}_R^2 \sim F_{(N_M - K_M, N_R - K_R)}$

Transformed model for $\mathrm{var}(e_i) = \sigma_i^2 = \sigma^2 x_i$

$y_i/\sqrt{x_i} = \beta_1(1/\sqrt{x_i}) + \beta_2(x_i/\sqrt{x_i}) + e_i/\sqrt{x_i}$

Estimating the variance function

$\ln(\hat{e}_i^2) = \ln(\sigma_i^2) + v_i = \alpha_1 + \alpha_2 z_{i2} + \cdots + \alpha_S z_{iS} + v_i$

Grouped data

$$\mathrm{var}(e_i) = \sigma_i^2 = \begin{cases} \sigma_M^2 & i = 1, 2, \ldots, N_M \\ \sigma_R^2 & i = 1, 2, \ldots, N_R \end{cases}$$

Transformed model for feasible generalized least squares

$$y_i \Big/ \sqrt{\hat{\sigma}_i} = \beta_1\left(1 \Big/ \sqrt{\hat{\sigma}_i}\right) + \beta_2\left(x_i \Big/ \sqrt{\hat{\sigma}_i}\right) + e_i \Big/ \sqrt{\hat{\sigma}_i}$$

Regression with Stationary Time Series Variables

Finite distributed lag model

$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_q x_{t-q} + v_t$

Correlogram

$r_k = \sum(y_t - \bar{y})(y_{t-k} - \bar{y}) / \sum(y_t - \bar{y})^2$

For $H_0: \rho_k = 0, \quad z = \sqrt{T} r_k \sim N(0, 1)$

LM test

$y_t = \beta_1 + \beta_2 x_t + \rho\hat{e}_{t-1} + \hat{v}_t \qquad$ Test $H_0: \rho = 0$ with t-test

$\hat{e}_t = \gamma_1 + \gamma_2 x_t + \rho\hat{e}_{t-1} + \hat{v}_t \qquad$ Test using $LM = T \times R^2$

AR(1) error $\quad y_t = \beta_1 + \beta_2 x_t + e_t \quad e_t = \rho e_{t-1} + v_t$

Nonlinear least squares estimation

$$y_t = \beta_1(1 - \rho) + \beta_2 x_t + \rho y_{t-1} - \beta_2 \rho x_{t-1} + v_t$$

ARDL(p, q) model

$$y_t = \delta + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + v_t$$

AR(p) forecasting model

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + v_t$$

Exponential smoothing $\quad \hat{y}_t = \alpha y_{t-1} + (1 - \alpha)\hat{y}_{t-1}$

Multiplier analysis

$$\delta_0 + \delta_1 L + \delta_2 L^2 + \cdots + \delta_q L^q = (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_p L^p) \times (\beta_0 + \beta_1 L + \beta_2 L^2 + \cdots)$$

Unit Roots and Cointegration

Unit Root Test for Stationarity: Null hypothesis:

$H_0: \gamma = 0$

Dickey-Fuller Test 1 (no constant and no trend):

$\Delta y_t = \gamma y_{t-1} + v_t$

Dickey-Fuller Test 2 (with constant but no trend):

$\Delta y_t = \alpha + \gamma y_{t-1} + v_t$

Dickey-Fuller Test 3 (with constant and with trend):

$\Delta y_t = \alpha + \gamma y_{t-1} + \lambda t + v_t$

Augmented Dickey-Fuller Tests:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{s=1}^{m} a_s \Delta y_{t-s} + v_t$$

Test for cointegration

$\Delta\hat{e}_t = \gamma\hat{e}_{t-1} + v_t$

Random walk: $\quad y_t = y_{t-1} + v_t$

Random walk with drift: $\quad y_t = \alpha + y_{t-1} + v_t$

Random walk model with drift and time trend:

$y_t = \alpha + \delta t + y_{t-1} + v_t$

Panel Data

Pooled least squares regression

$y_{it} = \beta_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + e_{it}$

Cluster robust standard errors $\quad \mathrm{cov}(e_{it}, e_{is}) = \psi_{ts}$

Fixed effects model

$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + e_{it} \qquad \beta_{1i}$ not random

$y_{it} - \bar{y}_i = \beta_2(x_{2it} - \bar{x}_{2i}) + \beta_3(x_{3it} - \bar{x}_{3i}) + (e_{it} - \bar{e}_i)$

Random effects model

$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + e_{it} \qquad \beta_{it} = \bar{\beta}_1 + u_i$ random

$y_{it} - \alpha\bar{y}_i = \bar{\beta}_1(1 - \alpha) + \beta_2(x_{2it} - \alpha\bar{x}_{2i}) + \beta_3(x_{3it} - \alpha\bar{x}_{3i}) + v_{it}^*$

$$\alpha = 1 - \sigma_e \Big/ \sqrt{T\sigma_u^2 + \sigma_e^2}$$

Hausman test

$$t = (b_{FE,k} - b_{RE,k}) \Big/ \left[\widehat{\mathrm{var}(b_{FE,k})} - \widehat{\mathrm{var}(b_{RE,k})}\right]^{1/2}$$

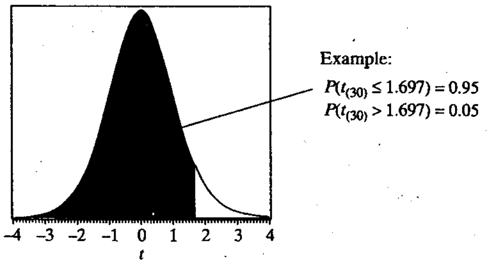
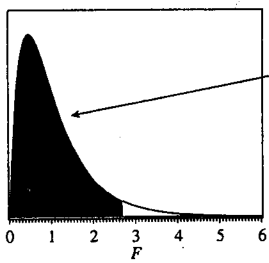


Table 2 **Percentiles of the *t*-distribution**

df	$t_{(0.90,\mathrm{df})}$	$t_{(0.95,\mathrm{df})}$	$t_{(0.975,\mathrm{df})}$	$t_{(0.99,\mathrm{df})}$	$t_{(0.995,\mathrm{df})}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715
38	1.304	1.686	2.024	2.429	2.712
39	1.304	1.685	2.023	2.426	2.708
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
∞	1.282	1.645	1.960	2.326	2.576

Source: This table was generated using the SAS® function TINV



Example:
$P(F_{(4,30)} \leq 2.69) = 0.95$
$P(F_{(4,30)} > 2.69) = 0.05$

Table 4 **95th Percentile for the F-distribution**

ν_2/ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	250.10	252.20	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.48	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.62	8.57	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.43	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.30	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.70	2.62	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.16	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.95	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.92	1.82	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.74	1.62
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.96	1.88	1.79	1.68	1.56
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.74	1.64	1.51
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.97	1.89	1.81	1.71	1.60	1.47
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.87	1.78	1.69	1.58	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.53	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.55	1.43	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.32	1.00

Source: This table was generated using the SAS® function FINV.