Name Mr. Key
ID 7777777

ECO 5350
Intro. Econometrics

Prof. T. Fomby
Spring 2011

Mid-Term Exam

Instructions: Put your name and student ID in the upper right-hand-corner of this exam. This exam is worth a total of 64 points. The breakout of these points by question is as follows:

Q1 = (3,3,3) = 9 points
Q2 = (2,2) = 4 points
Q3 = (2,2,2,2,2) = 10 points
Q4 = (3,4,4,3) = 14 points
Q5 = 5 points
Q6 = (10,2) = 12 points
Q7 = (3,3,4) = 10 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. Briefly define the following terms:

(3) a. homoscedasticity The error variance in the regression model is invariant to any choice of the independent (explanatory) variables. $E(e_i^2) = \sigma^2$ for all $i = 1, 2, \ldots, N$.

(3) b. BLUE stands for Best Linear Unbiased Estimator. The estimator that has the smallest sampling variance among linear, unbiased estimators.

(3) c. Gauss-Markov Theorem Among the class of linear unbiased estimators, the Least Squares estimators have minimum variance.

2. Let E(x) = 2, Var(x) = 4, E(y) = 1, Var(y) = 6, and Cov(x,y) = 2. Compute the following quantities. Show your work for full credit.

(2) $E(4x + 5y) = 4E(x) + 5E(y) = 4 \cdot 2 + 5 \cdot 1 = 13$

(2) $Var(3x - 4y) = 3^2 Var(x) + 4^2 Var(y) - 2(3)(4)cov(x,y)$
$= 9 \cdot 4 + 16 \cdot 6 - 24 \cdot 2 = 84$

3. a. What was the purpose of discussing the Extreme Value Estimator $b_2^{(EV)}$ in class on Tuesday? To show that this linear, unbiased estimator, although linear and unbiased ~~is~~ is not efficient relative to the least squares estimator thus demonstrating the Gauss-Markov theorem.

(2)

b. The Extreme Value Estimator of the slope of the conditional mean function, i.e. β_2, is given by the formula

(2)
$$b_2^{(EV)} = \frac{Y_{(n)} - Y_{(1)}}{X_{(n)} - X_{(1)}}$$

note: $cov(Y_{(n)}, Y_{(1)}) = 0$.

c. The variance of this estimator is given by the formula

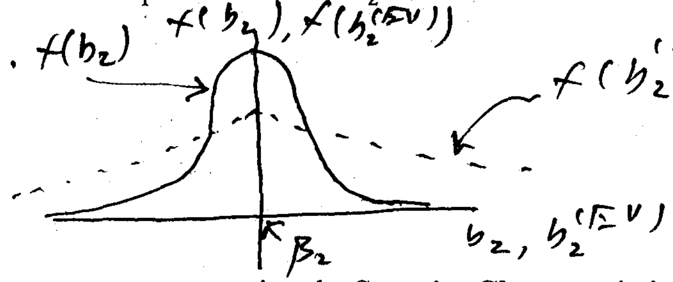
(2)
$$Var(b_2^{(EV)}) = \frac{1}{[X_{(n)} - X_{(1)}]^2} \cdot Var(Y_{(n)}) + Var(Y_{(1)})$$
$$= \frac{2\sigma^2}{[X_{(n)} - X_{(1)}]^2} > var(b_2)$$

d. True or False. $E\left(b_2^{(EV)}\right) = \beta_2$, $Var\left(b_2^{(EV)}\right) > Var(b_2)$ and $b_2^{(EV)}$ is a linear estimator of β_2.

(2)

(2) e. Illustrate the Gauss-Markov Theorem by drawing in the same graph the sampling distributions of the least squares estimator b_2 and the Extreme Value estimator $b_2^{(EV)}$.

sampling dist. $f(b_2)$ of b_2



$f(b_2^{(EV)})$ = sampling distribution of $b_2^{(EV)}$

b_2 has smaller sampling var. than $b_2^{(EV)}$ thus demonstrating Gauss-Markov thm.

4. This question covers concepts concerning the **Security Characteristic Line** and **Jensen's alpha** in financial economics. In this case y = msft– rkfree, x = mkt – rkfree where msft stands for monthly returns to Microsoft 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: 03/22/11 Time: 22:07
Sample: 1998M01 2008M12
Included observations: 132

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.013737	0.009061	1.516086	0.1319
X	1.259919	0.156861	8.032071	0.0000

R-squared	0.331668	Mean dependent var	-0.023547
Adjusted R-squared	0.326527	S.D. dependent var	0.108938
S.E. of regression	0.089400	Akaike info criterion	-1.976347
Sum squared resid	1.039016	Schwarz criterion	-1.932669
Log likelihood	132.4389	Hannan-Quinn criter.	-1.958598
F-statistic	64.51416	Durbin-Watson stat	2.348050
Prob(F-statistic)	0.000000		

(3) a. Jensen's alpha $\hat{\alpha}$ in this case is <u>0.013737</u>. Its standard error is <u>0.009061</u>. Its t-statistic is <u>1.516086</u>

(4) b. Compute a **90%** confidence interval for Jensen's alpha. It is [<u>-0.001163</u>, <u>0.028637</u>]. Show your work below.

$C.I.: b_1 \pm se(b_1) t_c \Rightarrow 0.013737 \pm (.009061)(1.645)$

$t_c = t_{\infty,.95} = 1.645$

$\Rightarrow 0.013737 \pm 0.01490 = [-0.001163, 0.028637]$

c. Over the time the Microsoft stock is observed, did it return a superior risk-adjusted rate of return according to Jensen's alpha? Use a **10 % level of significance** and appropriate null and alternative hypotheses to make your point.

(4)

$H_0: \beta_1 = 0$

$H_1: \beta_1 > 0$

we want a one-tailed p-value in this case and it is $0.1319/2 = 0.06595$. This indicates statistical significance at the 10% level.

$t_{\infty,.90} = 1.282$

observed $t = 1.576 > 1.282$

Therefore, we conclude that Microsoft stock produced a superior risk-adjusted return.

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

(3) Since its "beta" = 1.259919 and this is greater than 1.0, we judge Microsoft stock to be an aggressive stock as it is more volatile than the S&P 500 index.

5. Consider the following two regressions obtained using the R.C. Fair data on presidential voting.

Dependent Variable: VOTE
Method: Least Squares
Date: 03/22/11 Time: 22:29
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51.88113	0.951330	54.53536	0.0000
GROWTH	0.639044	0.162055	3.943371	0.0005
WAR	-3.052908	3.256669	-0.937433	0.3563
PARTY	-0.639303	0.941823	-0.678793	0.5027

R-squared	0.391683	Mean dependent var	52.09939
Adjusted R-squared	0.328753	S.D. dependent var	6.056635
S.E. of regression	4.962180	Akaike info criterion	6.154780
Sum squared resid	714.0737	Schwarz criterion	6.336175
Log likelihood	-97.55387	Hannan-Quinn criter.	6.215814
F-statistic	6.224162	Durbin-Watson stat	2.286111
Prob(F-statistic)	0.002142		

Dependent Variable: VOTE
Method: Least Squares
Date: 03/22/11 Time: 22:33
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51.69080	0.871102	59.33955	0.0000
GROWTH	0.654512	0.161080	4.063277	0.0003

R-squared	0.347509	Mean dependent var	52.09939
Adjusted R-squared	0.326461	S.D. dependent var	6.056635
S.E. of regression	4.970647	Akaike info criterion	6.103669
Sum squared resid	765.9272	Schwarz criterion	6.194366
Log likelihood	-98.71053	Hannan-Quinn criter.	6.134186
F-statistic	16.51022	Durbin-Watson stat	2.226791
Prob(F-statistic)	0.000306		

James Carville, a famous advisor of former President Bill Clinton, established the slogan of Clinton's first campaign: "It is the economy stupid." Let's test Carville's hypothesis.

Conclusion: Wars and Political Affiliation of Incumbent party does not matter. ✓

Let us test, via a Wald F-statistic, the null hypothesis that growth in the economy is the only thing that matters and that whether the country is at war before the election or whether the incumbent party is Republican or Democrat does not matter. Test this proposition at the **5% level of significance**. Be sure and clearly establish the acceptance and rejection regions of your test and clearly state the conclusion your draw from your test.

(5)

$H_0: \beta_3 = \beta_4 = 0$ versus H_1: not H_0.

$$F = \frac{(SSE_r - SSE_u)/J}{SSE_u/(N-K)} = \frac{(765.9272 - 714.0737)/2}{714.0737/(33-4)} = \frac{25.926}{24.623} = 1.05$$

$F_c = 3.32$ $1.05 < 3.32$ Accept H_0.

Wars and Party don't matter.

6. a. Use the **first regression output of question** 5 to fill in the following ANOVA table.

(10)

from output

Source	SS	DF	MS	F(overall)
Model	459.7775	3	153.259	6.22 (p= .002)
Error	714.0737	29	24.623	
Total	1173.8513	32		

from output K−1 N−K N−1 from output

from output

$R^2 = 0.391683$ $N = 33$, $K = 4$

Show your work to receive full credit.

$$1 - R^2 = \frac{SSE}{TSS} \therefore TSS = \frac{SSE}{1-R^2}$$

$$TSS = \frac{714.0737}{1 - 0.391683} = 1173.8513 \quad ; \quad TSS = SSE + SSM$$

$$SSM = TSS - SSE = 1173.8513 - 714.0737 = 459.7775$$

b. What is the null hypothesis that the above table is testing?

(2)

$H_0: \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_1: not H_0.

That is Growth, War, and Party are, jointly, not statistically significant.

7. Consider the following two regression outputs to answer the following questions: These regressions refer to 1080 home prices in Baton Rouge, Louisiana in 2010. The variable PRICE is the price in dollars of the home, SQFT is the number of square feet in the home, BATHS is the number of bathrooms in the home, and AGE is the age in years of the home. The other variables are defined as follows: SQFT_S = SQFT – 3500, BATHS_S = BATHS – 3, and AGE_S = AGE – 14.

Dependent Variable: PRICE
Method: Least Squares
Date: 03/22/11 Time: 23:03
Sample: 1 1080
Included observations: 1080

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-82750.42	9490.954	-8.718873	0.0000
SQFT	75.98158	3.371207	22.53839	0.0000
BATHS	35840.63	5751.989	6.230998	0.0000
AGE	-502.3845	144.2971	-3.481599	0.0005

R-squared	0.603885	Mean dependent var	154863.2
Adjusted R-squared	0.602780	S.D. dependent var	122912.8
S.E. of regression	77466.26	Akaike info criterion	25.35677
Sum squared resid	6.46E+12	Schwarz criterion	25.37523
Log likelihood	-13688.66	Hannan-Quinn criter.	25.36376
F-statistic	546.7935	Durbin-Watson stat	1.900395
Prob(F-statistic)	0.000000		

Dependent Variable: PRICE
Method: Least Squares
Date: 03/22/11 Time: 23:14
Sample: 1 1080
Included observations: 1080

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	283673.6	4627.094	61.30708	0.0000
SQFT_S	75.98158	3.371207	22.53839	0.0000
BATHS_S	35840.63	5751.989	6.230998	0.0000
AGE_S	-502.3845	144.2971	-3.481599	0.0005

R-squared	0.603885	Mean dependent var	154863.2
Adjusted R-squared	0.602780	S.D. dependent var	122912.8
S.E. of regression	77466.26	Akaike info criterion	25.35677
Sum squared resid	6.46E+12	Schwarz criterion	25.37523
Log likelihood	-13688.66	Hannan-Quinn criter.	25.36376
F-statistic	546.7935	Durbin-Watson stat	1.900395
Prob(F-statistic)	0.000000		

a. Do the variables in the Baton Rouge house price equation significantly explain the variation in house prices in Baton Rouge? Explain your answer.

3

Yes. The overall F-statistic 546.7935 has a p-value of 0.0000 and thus is highly significant as the p-value is less than 0.05.

b. Suppose that I am interested in predicting the price of a home that I am thinking about purchasing in Baton Rouge. Suppose the house has 3500 square feet of space, 3 bathrooms, and is 14 years of age. What is your BLU estimate of the price you might expect to see the house eventually sell for (in dollars). using the second transformed regression, $\hat{\theta}_1 = \$283,673.60$ is the predicted price. This is a BLU estimate as it is a linear combination of the BLU estimates b_2, b_3, and b_4.

3

c. Although you don't know what the price of the house will be when it is finally sold, construct a 95% confidence interval for the eventual sales price of the home. Show you work if you expect full credit. The 95% confidence interval is [131,569, 435,778].

4

$$se(\hat{y}_0) = \sqrt{se(\hat{\theta}_1)^2 + (SER)^2}$$

$$= \sqrt{(4627.094)^2 + (77466.26)^2}$$

$$= 77,627.094$$

95% C.I.:

$$\hat{\theta}_1 \pm se(\hat{y}_0) t_c \quad , \quad t_c = t_{\infty, .975} = 1.96$$

$$283,673.60 \pm 77,627.094(1.96)$$

$$283,673.60 \pm 152,104.48$$

$$(131,569.12; 435,778.0793)$$

The Rules of Summation

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

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

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

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

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

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

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

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

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

Expected Values & Variances

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

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

$E[g_1(X) + g_2(X)] = \sum_x [g_1(x) + g_2(x)] f(x)$
$= \sum_x g_1(x) f(x) + \sum_x g_2(x) f(x)$
$= E[g_1(X)] + E[g_2(X)]$

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

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

$\operatorname{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$
$= \sum_x\sum_y [x - E(X)][y - E(Y)] f(x, y)$

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

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

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

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

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

Normal Probabilities

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

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 $\operatorname{var}(e) = \sigma^2 = \operatorname{var}(y)$

SR4 The covariance between any pair of random errors, e_i and e_j is $\operatorname{cov}(e_i, e_j) = \operatorname{cov}(y_i, y_j) = 0$

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

t-test for More than One Parameter

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

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

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

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

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

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

Collinearity and Omitted Variables

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

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

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

Heteroskedasticity

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

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 $\text{var}(e_i) = \sigma_i^2 = \sigma^2 x_i$

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

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

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

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

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

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

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

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

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: $y_t = y_{t-1} + v_t$

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

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 $\text{cov}(e_{it}, e_{is}) = \psi_{ts}$

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

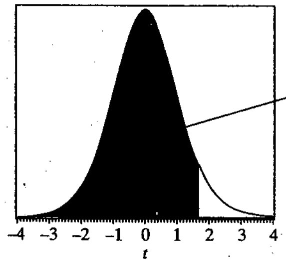
$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + e_{it} \qquad \beta_{it} = \bar{\beta}_1 + u_i$ random

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

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

Hausman test

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



Example:
$P(t_{(30)} \le 1.697) = 0.95$
$P(t_{(30)} > 1.697) = 0.05$

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

df	$t_{(0.90,df)}$	$t_{(0.95,df)}$	$t_{(0.975,df)}$	$t_{(0.99,df)}$	$t_{(0.995,df)}$
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

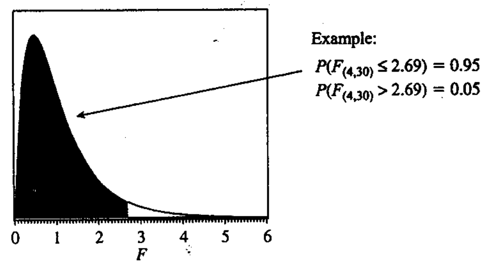


Table 4 95th Percentile for the F-distribution

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.