A Typical Regression Analysis

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

53

General Info

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

54

Parameter Estimates

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

55

Parameter Estimates

Variable	Coefficient
C	9.884732
X	1.073140
Z	-0.638011

The “coefficients" are simply the regression coefficient estimates. Our estimated regression equation is

$$\hat{y} = 9.884732 + 1.073140x - 0.638011z$$

56

Standard Errors of Parameter Estimates

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

57

Standard Errors of Parameter Estimates

Std. Error
0.190297
0.150341
0.172499

The standard errors of the estimated coefficients indicate their sampling variability, and hence their reliability. The estimated coefficient plus or minus one standard error is approximately a 68% confidence interval for the true but unknown population parameter, and the estimated coefficient plus or minus two standard errors is approximately a 95% confidence interval, assuming that the estimated coefficient is approximately normally distributed.4 Thus large coefficient standard errors translate into wide confidence intervals.

58

Significance Test Statistics: H_0: Parameter = 0

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

59

Significance Test Statistics: H_0: Parameter = 0

t-Statistic
51.94359
7.138031
-3.698642

Each t statistic provides a test of the hypothesis variable irrelevance: that the true but unknown population parameter is zero, so that the corresponding variable contributes nothing and can therefore be dropped. The t statistic is just the ratio of the estimated coefficient to its standard error.

60

Significance Test Statistics: P-values

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

61

Significance Test Statistics: P-values

Prob.
0.0000
0.0000
0.0006

Associated with each t statistic is a probability value, or p-value, which is the probability of getting a value of the t statistic at least as large in absolute value as the one actually obtained, assuming that the null hypothesis (irrelevance of the variable) is true. The smaller the p-value, the stronger the evidence against irrelevance.

62

Other Statistical Info

Table 1
Regression of y on x and z

LS // Dependent Variable is Y

Sample: 1960 2007
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.884732	0.190297	51.94359	0.0000
X	1.073140	0.150341	7.138031	0.0000
Z	-0.638011	0.172499	-3.698642	0.0006

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

63

R-squared

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

A standard measure of overall goodness of fit is R2 (R-squared), technically the percentage of the variance of y explained by the variables in the model:

$$R^2 = \frac{SSModel}{SSTotal} = 1 - \frac{SSE}{SSTotal}$$

64

Adjusted R^2

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

Adjusts for the number of predictors in the model, penalizing (reducing the value) for each additional variable. A model with no "irrelevant" variables will have an adjusted R^2 almost as large as R^2.

$$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SSTotal/(n-1)} = 1 - \frac{MSE}{MST}$$

65

S.E. of Regression

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is $\sqrt{MSE}$. It is an unbiased estimator of σ, the standard deviation of the random errors (noise).

66

Sum Squared Residuals

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is SSE, the quantity we minimize to find the least-squares regression model.

67

Log Likelihood

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

The likelihood function is tremendously important in statistics, as it summarizes all the information contained in the data. It is simply the joint density function of the data, viewed as a function of the model parameters. A great method of optimization is called maximum likelihood estimation, and LS estimators are also maximum likelihood estimators. The given value is the maximum value of the likelihood function. It can be useful when comparing models, but on its own is kind of useless.

68

Durbin-Watson Stat

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

The Durbin-Watson statistic tests lag 1 serial correlation; i.e., is the last residual correlated to the present residual. In time series, we check for serial correlation at ALL lags, not just lag 1.

If there is no serial correlation at the tested lag, the DW stat should have a value of 2 (it is always between 0 and 4). Values less than 2 indicate positive correlation, and values greater than 2 indicate negative correlation.

69

Mean Dependent Variable

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is just the value of $\bar{y}$.

70

S.D. Dependent Variable

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is just the value of s_y^2.

71

Akaike Info Criterion (AIC) and Schwarz criterion (SIC)

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

The Akaike and Schwarz criteria are used for model selection, and in certain contexts they have provable optimality properties in that regard. Both are penalized versions of MSE, where the penalties are functions of the degrees of freedom used in fitting the model. For both AIC and SIC, “smaller is better."

SIC is also called Bayes Info Criterion, or BIC.

72

F-statistic

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

We use the F statistic to test the hypothesis that the coefficients of all variables in the regression except the intercept are jointly zero, versus the alternative that at least one of the coefficients is not zero.

73

F-statistic

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

The F-statistic is computed from values sometimes given in an ANOVA table.

Source	d.f.	Sum of Squares	Mean Square	t.s.	p-value
Model	***K***	***SSRegression***	***SSR/K***	$\frac{MSR}{MSE}$	$F_{K,T-K-1}$
Error	***T-K-1***	***SSE***	***SSE/(T-K-1)***		
Total	***T-1***	***SSTotal***			

74

Prob(F-statistic)

R-squared	0.552928	Mean dependent var	10.08241
Adjusted R-squared	0.533059	S.D. dependent var	1.908842
S.E. of regression	1.304371	Akaike info criterion	3.429780
Sum squared resid	76.56223	Schwarz criterion	3.546730
Log likelihood	-79.31472	F-statistic	27.82752
Durbin-Watson stat	1.506278	Prob(F-statistic)	0.000000

This is the p-value associated with the F-statistic.

In the case of only one right-hand-side variable, the t and F statistics contain exactly the same information, and one can show that $F = t^2$.

75

3.4 Regression From a Forecasting Perspective

76

The Key to Everything (or at Least Many Things)

Linear least squares regression, by construction, is consistent under very general conditions for "the linear function of x_t that gives the best approximation to y_t under squared-error loss."

This is a projection (like your shadow on a sunny day).

We project "reality" (the data) onto our model.

77

The Key to Everything (or at Least Many Things)

To forecast y_t for any given value of x_t, we can use the fitted line to find the value of y_t that corresponds to the given value of x_t.

This is the optimal forecast under quadratic loss, or the best linear approximation to it.

78

Why Take a Probabilistic Approach to Regression, as Opposed to Pure Curve Fitting?

We may want to test hypotheses regarding which variables.

We want to quantify the uncertainty associated with our forecasts (that is, we want interval and density forecasts).

79

Why Take a Probabilistic Approach to Regression, as Opposed to Pure Curve Fitting?

The regression function is the conditional expectation of y_t given x_t.

This is crucial for forecasting, because the expectation of future y conditional upon available information is a particularly good forecast.

In fact, under quadradic loss, it is the best possible forecast.

80

Residual Plots

Residual plots are useful for visually flagging neglected things that impact forecasting.

Residual serial correlation indicates that point forecasts could be improved.

81

Chapter 4

We will skip Chapter 4. It is mainly concerned with high-level regression techniques to address specific problems in the FIC. These techniques are highly problem specific, and require a greater mathematical statistics understanding of the multiple regression model than is feasible in this course.

82