Ordinary Least Squares (OLS) Regression

CAPM

$E(R) = R_f + \beta[E(R_M) - R_f]$

$E(R) = R_f + \beta[E(R_M) - R_f] + \varepsilon$

What is a good estimator?

1) Computational cost

2) Least squares

3) Highest R^2

4) Unbiasedness

Regression versus Causation

Regression versus Correlation

Types of regression

1) Time series

2) Cross sectional

3) Pooled

OLS Assumptions

1) Dependent variable is a linear function of the independent variable(s).

Specification error
a) Wrong regress – omitted or irrelevant
b) nonlinearity
c) changing parameters

2) X values are fixed in repeated samples

3) Expected value of disturbance error is zero.

4) Homoscedasticity or equal variances if Xs

Heteroscedasticity

5) No autocorrelation of errors

6) No covariance between X and ε

7) Number of observations must be greater than the number of Xs.

8) Variability in X

9) Regression is correctly specified. No specification error.

10) No multicollinearity. No perfect linear relationship among Xs.

R^2

$\sum(Y_i - \bar{Y})^2$ = Total sum of squares (TSS)

$\sum \varepsilon_i^2$ = Residual sum of squares (RSS)

$\sum(\hat{Y}_i - \bar{Y})^2$ = Explained sum of squares (TSS)

$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$

$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$

R^2 has to be non-negative

$0 \leq R^2 \leq 1$

Adjusted R^2 → Adjust for the number of regressors

Adjusted $R^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k}\right)$

Where:

n = number of observations

k = number of parameters including intercept

F statistic

F tests if all slope coefficients are simultaneously equal to zero

CAPM to the Market Model

$E(R) = R_f + \beta[E(R_M) - R_f]$

$E(R) - R_f = \beta[E(R_M) - R_f]$

$E(R) - R_f = \alpha + \beta[E(R_M) - R_f] + \varepsilon$