/02 Hypothesis Tests and Confidence Intervals In Multiple Regression

Hypothesis Tests and Confidence Intervals In Multiple Regression

- **Hypothesis Testing of Regression Coefficients:**
  - ❑In a multiple regressions, the magnitude of the coefficients tells us nothing about the importance of the independent variable in explaining the dependent variable.
  - ❑Thus, we must conduct hypothesis testing on the estimated slope coefficients to determine if the independent variables make a significant contribution to explaining the variation in the dependent variable.
  - ❑The t-statistic used to test the significance of the individual coefficients in a multiple regression.

$$t = \frac{b_j - B_j}{s_{b_j}} = \frac{\text{estimated regression cofficient} - \text{hypothesized value}}{\text{coefficient standard error of } b_j}$$

, where df=n-k-1

  - ❑"testing statistical significance" –> $H_0 : b_j = 0$ versus $H_A : b_j \neq 0$

Hypothesis Tests and Confidence Intervals In Multiple Regression

Hypothesis Testing of Regression Coefficients:

Example: Testing the statistical significance of a regression coefficient

Test the statistical significance of the independent variable PR in the real earnings growth example, at the 10% significance level. Assume that the number of observations is 46.

Coefficient and Standard error Estimates for Regression of EG10 on PR and YCS

	Coefficient	*Standard Error*
Intercept	-11.60%	1.66%
PR	0.25	0.032
YCS	0.14	0.28

Hypothesis Tests and Confidence Intervals In Multiple Regression

Hypothesis Testing of Regression Coefficients:

- **Interpreting p-Values :** The p-value is the smallest level of significance for which the null hypothesis can be rejected.
  - If the p-value is less than significance level, the null hypothesis can be rejected.
  - If the p-value is greater than the significance level, the null hypothesis cannot be rejected.

Example: Interpreting p-values

Given the following regression results, determine which regression parameters for the independent variables are statistically significantly different from zero at the 1% significance level, assuming the sample size is 60.

Variable	*Coefficient*	*Standard Error*	*t-Statistic*	*p-Value*
Intercept	0.4	0.4	1	0.3215
X1	8.2	2.05	4	0.0002
X2	0.4	0.18	2.2	0.0319
X3	-1.8	0.56	-3.2	0.0022

	Coefficient	*Standard Error*	*t-statistic*	*p-values*
Intercept	-11.60%	1.657%	-7	<0.0001
PR	0.25	0.032	7.8	<0.0001
YCS	0.14	0.280	0.5	0.62.

Hypothesis Tests and Confidence Intervals In Multiple Regression

➢ **Hypothesis Testing of Regression Coefficients:**

Example: Testing regression coefficients (two-tail test)

Using the previous data, test the null hypothesis that PR is equal to 0.20 versus the alternative that it is not equal to 0.20 using a 5% significance level.

Example: Testing regression coefficients (one-tail test)

Test the null hypothesis that the intercept term is greater than or equal to -10% versus the alternative that it is less than -10% using a 1% significance level.

Hypothesis Tests and Confidence Intervals In Multiple Regression

- **Hypothesis Testing of Regression Coefficients:**
  - **Confidence Intervals for a Regression Coefficient**

$$b_j \pm (t_c x s_{b_j})$$

Estimated regression coefficient (critical t-value)(coefficient standard error)

Example: Calculating a confidence intervals for a regression coefficient
Calculate the 90% confidence interval for the estimated coefficient for the independent variable PR in the real earnings Growth example.

Hypothesis Tests and Confidence Intervals In Multiple Regression

- **PREDICTING THE DEPENDENT VARIABLE:**

The predicate value of dependent variable Y is:

$$\hat{Y}_i = b_0 + b_1\hat{X}_{1i} + b_2\hat{X}_{2i} + \cdots \ldots + b_k\hat{X}_{ki}$$

Where:

$\hat{Y}_i$ =the predicted value of the dependent variable
B_j =the estimated slope coefficient for the jth independent variable
$\hat{X}_{ji}$=the forecast of the jth independent variable , j=1, 2, … ,k

Example: Calculate a predicated value for the dependent variable

An analyst would like to estimated regression from the previous example to calculate the predicted 10-year real earnings growth for the S&P 500, assuming the payout ratio of the index is 50%. He observes that the slope of the yield curve is currently 4%.

JOINT HYPOTHESIS TESTING:

In case of multiple variable, even if just one of the equality in hypothesis test doesn't hold, we reject the entire null hypothesis. Hence joint hypothesis is preferred where independent variables are correlated.

THE F-STATISTIC

- An F-test assesses how well the set of independent variables, as a group, explains the variation in the dependent variable.
- F-statistic is used to test whether at least one of the independent variable explains a significant portion of the variation of the dependent variable.

if there are four independent variables in the model, the hypothesis are structured as:

$H_0 : B_1 = B_2 = B_3 = B_4 = 0$ versus H_A: at least one $B_j \neq 0$.

The F-statistic, which is always a one-tailed test, is calculated as:

$$\frac{\frac{ESS}{k}}{\frac{SSR}{n-k-1}}$$

Where: ESS=explained sum of squares
SSR=sum of squared residuals
df (numerator)=k
df (denominator)=n-k-1

Decision rule: reject H_0 if F (test-statistic) > F_c (critical value)

Hypothesis Tests and Confidence Intervals In Multiple Regression

- **JOINT HYPOTHESIS TESTING:**

Example: Calculating and interpreting the F-statistic

An analyst runs a regression of monthly value-stock return on five independent variables over 60 months. The total sum of squares is 460, and the sum of squared residuals is 170. Test the null hypothesis at the 5% significance level (95% confidence) that all five of the independent variables are equal to zero.

Hypothesis Tests and Confidence Intervals In Multiple Regression

JOINT HYPOTHESIS TESTING:

INTERPRETING REGRESSION RESULTS

- The coefficient of determination is:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{SSR}{TSS} = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$$

- The coefficient of multiple correlation is the square root of R-squared.
- In the case of a multiple regression, the coefficient of multiple correlation is always positive.

Hypothesis Tests and Confidence Intervals In Multiple Regression

JOINT HYPOTHESIS TESTING:

INTERPRETING REGRESSION RESULTS

Example : ANOVA Table

$$\hat{Y}_i = -4.451 + 2.057 \times X_{1i} + 2.008 \times X_{2i}$$

R – squared	0.934					
Adj R-squared	0.890.					
Standard error	1.407					
Observations	6					
	Degrees of Freedom	*SS*	*MS*	*F*		
Explained	2	84.057	42.029	21.217		
Residual	3	5.943	1.981			
Total	5	90				
Variables	*Coeff.*	*Std. Error*	*t-stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-4.4511	3.299	-1.349	0.27	-14.95	6.048
Lockup	2.057	0.337	6.103	0.009	0.984	3.13
Experience	2.008	0.754	2.664	0.076	-0.391	4.407

The ANOVA table outputs the standard errors, t-statistics, probability values (p-values), and confidence intervals for the estimated coefficients.

Hypothesis Tests and Confidence Intervals In Multiple Regression

JOINT HYPOTHESIS TESTING:

SPECIFICATION BIAS

❑Specification bias refers to how the slope coefficient and other statistics for a given independent variable are usually different in a simple regression when compared to those of the same variable when included in a multiple regression. Example:

$\hat{Y}_i$ = 1+ 2 x $(\text{lockup})_i$
t =3.742

$\hat{Y}_i$ = 11.714 + 1.714 x $(\text{experience})_i$
t =2.386

$\hat{Y}_i$ =-4.451 + 2.057 x(lockup)I +2.008 x x $(\text{experience})_i$
t =6.103 t = 2.664.

Hypothesis Tests and Confidence Intervals In Multiple Regression

JOINT HYPOTHESIS TESTING:

R^2 AND ADJUSTED R^2

- The proportion of total variation in Y is explained by all X variables taken together.

$$R^2 = \frac{ESS}{TSS}$$

- R-squared never decreases when a new variable is added to the model.
- Adjusted R-squared
  - Its value depends on the number of explanatory variables.
  - Imposes a penalty for adding additional explanatory variables.

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k-1}(1-R^2)$$

Hypothesis Tests and Confidence Intervals In Multiple Regression

- **JOINT HYPOTHESIS TESTING:**

Restricted Vs. Unrestricted Least Square Models

Hypothesis Tests and Confidence Intervals In Multiple Regression

JOINT HYPOTHESIS TESTING:

MODEL MISSPECIFICATION

- Omitted variable bias in multiple regressions will result if the following two conditions occur:
  - The omitted variable is a determinant of the dependent variable.
  - The omitted variable is correlated with at least one of the independent variables.

Thanks
Leading Derivative & Risk Advisors

成蹊风险研究

riskmacro.com