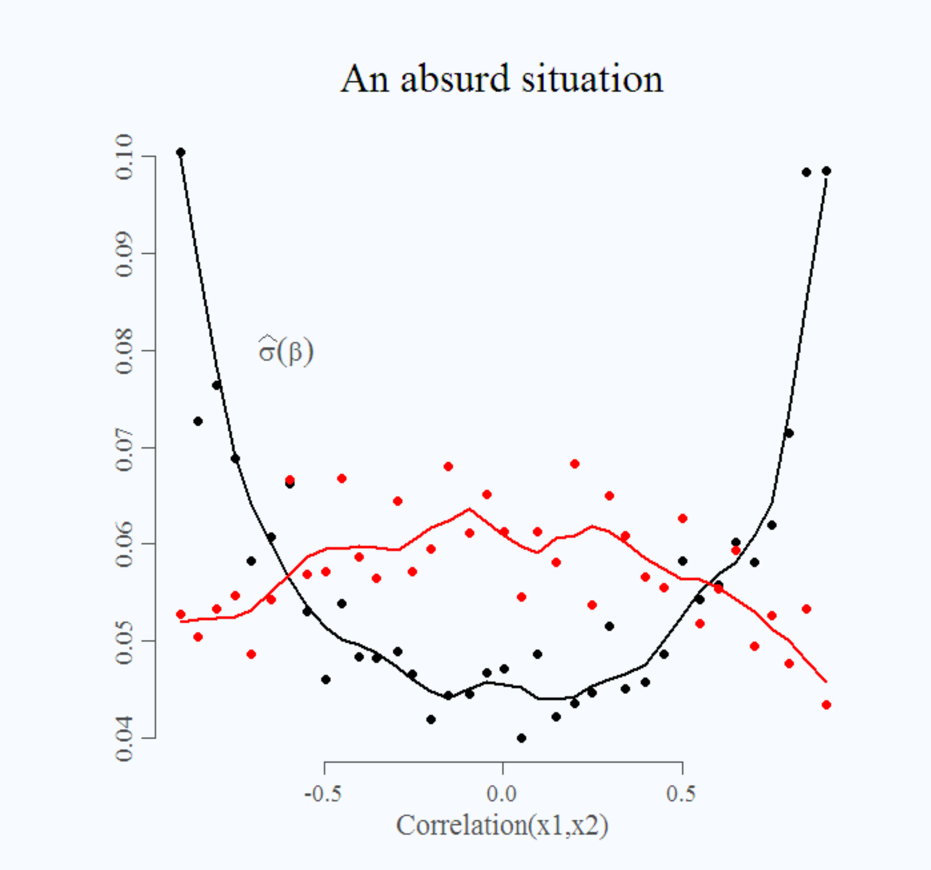
Linear Regression with Multiple Regressors

Linear Regression with Multiple Regressors

➢ **OMITTED VARIABLES BIAS:**

- ❑Omitting relevant factors from an ordinary least squares (OLS) regression can produce misleading or biased results.
- ❑Omitted variable bias is present when two conditions are met:
  1. The omitted variable is correlated with the movement of the independent variables in the model, and
  2. The omitted variable is a determinant of the dependent variable.
- ❑The correlations between the omitted variable and the independent variable will determine the size of the bias.
- ❑If a bias is found, it can be addressed by dividing data into group and examining one factor at a time while holding other factors constant.
- ❑Multiple regression analysis is used to eliminate omitted variable bias since it can estimate the effect of one independent variable on the dependent variable while holding all other variables constant.



In black, the standard deviation when the model is correctly specified. The parabolic shape is due to multicollinearity. In red is the (estimated) standard deviation of the incorrectly specified model.

Linear Regression with Multiple Regressors

➢ **MULTIPLE REGRESSION BASICS :**

- ✓ **Multiple regression is** regression analysis with more than one independent variable.
- ✓ The general multiple linear regression model is:

$$Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \cdots B_k X_{ki} + \varepsilon_i$$

  - The intercept term is the value of the dependent variable when the independent variables are all equal to zero.
  - Each slope coefficient is the estimated change in the dependent variable for a one-unit change in that independent variable, holding the other independent variable constant. That’s why the slope coefficients in a multiple regression are sometimes called partial slope coefficients.

Linear Regression with Multiple Regressors

➢ **MULTIPLE REGRESSION BASICS :**

- The estimators of there coefficients are known as ordinary lest squares (OLS) estimators.
- The regression equation can be expressed as:

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

- The residual, e_i

$$e_i = Y_i - \hat{Y}_i$$

Linear Regression with Multiple Regressors

- **MULTIPLE REGRESSION BASICS :**

Example: Formulating the Multiple Regression Equation

The authors formulate the following regression equation using annual data (46 observations):

$$RG10 = B_0 + B_1PR + B_2YCS + \varepsilon$$

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

Linear Regression with Multiple Regressors

- **MULTIPLE REGRESSION BASICS :**

  Homoskedasticity refers to the condition that the variance of the error term is constant for all independent variables, X, from i=1 to n: $Var(\varepsilon_i / \mathbb{X}_i) = \sigma^2$

  Heteroskedasticity means the dispersion of the error terms varies over the sample.

  Conditional heteroskedasticity refers the variance is a function of independent variables.

Linear Regression with Multiple Regressors

MEASURES OF FIT:

- The standard error of the regression (SER) measures the uncertainty about the accuracy of the predicted values of the dependent variable.
- SER is the standard deviation of the predicted values for the dependent variable about the regression line.

$$SER = \sqrt{s_c^2} = \sqrt{\frac{SSR}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n}[Y_i - (b_0 + b_i X_i)]^2}{n-k-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-k-1}}$$

- SER measures the degree of variability of the actual Y-values relative to the estimated Y-values.
- The SER gauges the “fit” of the regression line. The smaller the standard error, the betted the fit.

Linear Regression with Multiple Regressors

- **COFFICIENT OF DETERMINATION, R^2:**

  - The multiple coefficient of determination, R^2, is used to test the overall effectiveness of the entire set of independent variables in explaining the dependent variable.

$$R^2 = \frac{total\ variation - unexplained\ variation}{total\ variatiion} = \frac{TSS - SSR}{TSS}$$

$$= \frac{explained\ variation}{total\ variation} = \frac{ESS}{TSS}$$

Linear Regression with Multiple Regressors

➢ COFFICIENT OF DETERMINATION, R^2:

❑ **Adjusted R^2**

- ○ R^2 by itself may not be a reliable measure of the explanatory power of the multiple regression model.
- ○ This is because R^2 almost always increase as independent are added to the model, even if the marginal contribution of the new variables is not statistically significant.
- ○ Consequently, a relatively high R^2 may reflect the impact of a large set of independent variables rather then how well the set explains the dependent variable.
- ○ This problem if often referred to as overestimating the regression.
- ○ To overcome the problem of overestimating the impact of additional variables on the explanatory power of a regression model, we use adjusted R^2.
- ○ The adjusted R^2 value is expressed as: $R_a^2 = 1 - \left[\left(\frac{n-1}{n-k-1}\right) x(1 - R^2)\right]$

Linear Regression with Multiple Regressors

➢ **COFFICIENT OF DETERMINATION, R^2:**

Example: Calculating R^2 and adjusted R^2

An analyst rens a regression of monthly value-stock returns on five independent variables over 60 months. The total sum of the regression is 460, and the sum of squared error is 170. Calculate the R^2 and adjusted R^2

Linear Regression with Multiple Regressors

- **COFFICIENT OF DETERMINATION, R^2:**

Example: Interpreting adjusted R^2

Suppose the analyst now adds four more independent variables to regression and the R^2 increase to 65.0%. indentify which model the analyst would most likely prefer.

Linear Regression with Multiple Regressors

- **ASSUMPTIONS OF MULTIPLE REGRESSION:**

  - Most of the assumptions made with the multiple regression pertain to ε, the model's error term:
    - A linear relationship exists between the dependent and independent variables.
    - The independent variables are not random, and there is no exact linear relation between any two or more independent variables.
    - The expected value of the error term, conditional on the independent variables, is zero.
    - The variance of the error term is constant for all observation.
    - The error term for one observation is not correlated with that of another observation.
    - The error term is normally distributed

Linear Regression with Multiple Regressors

- **MULTICOLLINEARITY:**

❑**Multicollinearity** refers to the condition when two or more of the independent variables, or linear combinations of the independent variables, in a multiple regression are highly correlated with each other.

❑Perfect multicollinearity arises if one of the independent variables is a perfect linear combination of the other independent variables.

❑Imperfect multicollinearity arises when two or more independent variables are highly correlated, but less than perfectly correlated.

Effect of Multicollinearity on Regression Analysis

There is a greater probability to conclude that a variable is not statistically significant (e.g., a Type II error).

Detecting Multicollinearity

- Check the situation where t-test indicates that none of the individual coefficients is significantly different than zero, while the R^2 is high.
- High correlation among independent variables also suggest as a sign of multicollinearity.
- But also low correlation among the independent variables does not necessarily indicate multicollinearity is not present.

Correcting Multicollinearity

- Omit one or more of the correlated independent variables.
- Use stepwise regression , which systematically remove variables from the regression until multicollinearity is minimized.

Linear Regression with Multiple Regressors

- **MULTICOLLINEARITY:**

Example: Detecting multicollinearity

Bob Watson runs a regression of mutual fund returns on average P/B, average P/E, and average market capitalization, with the following results:

Variable	*Coefficient*	*p-value*
Average P/B	3.52	0.15
Average P/E	2.78	0.21
Market Cap	4.03	0.11
R^2	89.60%	

Determine whether or not multicollinearity is a problem in this regression.

Thanks
Leading Derivative & Risk Advisors

成蹊风险研究

riskmacro.com