Quantitative Methods 2

Tutorial 10

Nhan La

Last week

- Multiple linear regression

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

- General F test
  - Compare unrestricted against restricted model
  - Test linear restriction of parameters
  - General expression: $F = \frac{n-k-1}{m}\frac{SSE_r - SSE}{SSE}$
  - Same dependent variable: $F = \frac{n-k-1}{m}\frac{R^2 - R_r^2}{1-R^2} = \frac{(R^2 - R_r^2)/m}{(1-R^2)/(n-k-1)}$
- Conduct the linear restriction test
  - Use EViews' coefficient restriction test, or
  - Estimate separately restricted and unrestricted models

Tutorial 10

MLR assumptions and violations

	Assumption	Violation
MLR1	$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$	Nonlinearity
MLR2	$E(\varepsilon_i \mid x_i) = 0$	Omitted variable
MLR3	$Var(\varepsilon_i \mid x_i) = \sigma^2$	Heteroskedasticity
MLR4	$E(\varepsilon_i \varepsilon_j \mid x_i, x_j) = 0$	Autocorrelation
MLR5	No exact linear relationship among x_i	(Perfect) multicollinearity
MLR6	$\varepsilon_i \mid x_i \sim N(0, \sigma^2)$	Unreliable hypothesis testing

Tutorial 10

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

- Multicollinearity
  - Effect of X_i on Y, holding all other X's constant?

- Heteroskedasticity : $\mathrm{Var}(\varepsilon_i | x_i) = \sigma_i^2$

- Consequence:
  - $\hat{\beta}$ are still linear and unbiased
  - $\mathrm{SE}(\hat{\beta})$ are incorrect, thus hypothesis tests are incorrect

Tutorial 10

Violation of MLR5: Multicollinearity

- High R^2 but only a few significant t-ratios with the logical signs
  - $t_k = \frac{\hat{\beta}_k - 0}{SE(\hat{\beta}_k)}$
  - $R^2 \geq 0.5$
- Strong correlation between X's. Stronger correlation between some X's than between these variables and Y
  - $|r| > 0.8$
- $VIF_j = \frac{1}{1-R_j^2} \geq 5$

Tutorial 10

Violation of MLR3: Heteroskedascity

$$\mathrm{Var}(\varepsilon_i|x_i) = \sigma_i^2$$

- OLS estimators $(\hat{\beta})$ are still linear, unbiased, consistent and normally distributed

- $\mathrm{Var}(\hat{\beta})$ is often underestimated
  - Inflate significance of the hypothesis testing

Tutorial 10

Violation of MLR3: Heteroskedascity

- Plot e_i against independent variables and observe patterns
- White test

i. $y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + e_i$

ii. Auxiliary model

$$e_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1i} + \hat{\alpha}_2 X_{2i} + \hat{\alpha}_3 X_{1i}^2 + \hat{\alpha}_4 X_{2i}^2 + \hat{\alpha}_5 X_{1i} X_{2i} + v_i$$

Why?

$$\mathrm{Var}(e_i|X_i) = E\left(e_i^2|X_i\right) - \underbrace{[E(e_i)|X_i]^2}_{0}$$

$$\mathrm{Var}(e_i|X_i) = E\left(e_i^2|X_i\right)$$

iii. Test statistic

$$W = nR_{aux}^2 \sim \chi_{df}^2$$

H_0: model is homoskedastic; H_A: model is heteroskedastic

df = number of regressors (independent variables) <u>in the auxiliary model</u>

n = sample size