Session 3: Model diagnostics and specification

Summer school

June 4, 2015

Testing the Stability

$$\hat{Y}_t = 62.4226 + 0.0376 X_t$$

t = (4.8917) (8.8937)

$R^2 = 0.7672$ $RSS_3 = 23{,}248.30$ df=24

1970-1981 $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}$ $n_1 = 12$

1982-1995 $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $n_2 = 14$

1970-1995 $Y_t = \alpha_1 + \alpha_2 X_t + u_t$ $n = 26$

$\alpha_1 = \gamma_1 = \lambda_1$ $\alpha_2 = \gamma_2 = \lambda_2$

Testing the Stability

$$\hat{Y}_t = 62.4226 + 0.0376X_t$$

t = (4.8917) (8.8937)

$R^2 = 0.7672$ $RSS_3 = 23{,}248.30$ df=24

$$\hat{Y}_t = 1.0161 + 0.0803X_t$$

t = (0.0873) (9.6015)

$R^2 = 0.9021$ $RSS_3 = 1785.032$ df=10

$$\hat{Y}_t = 153.4947 + 0.0148X_t$$

t = (4.6922) (1.7707)

$R^2 = 0.2971$ $RSS_3 = 10{,}005.22$ df=12

Testing the Stability: Chow Test

Assumptions

$$u_{1t} \approx NID(0, \sigma^2) \quad u_{2t} \approx NID(0, \sigma^2)$$

Mechanics

$$RSS_{UR} = RSS_1 + RSS_2 \qquad df = (n_1 + n_2 - 2k)$$

$$F = \frac{(RSS_R - RSS_{UR}) / k}{RSS_{UR} / (n_1 + n_2 - 2k)} \approx F_{[k,(n_1+n_2-2k)]}$$

$$F = \frac{(23{,}248.3 - 11{,}790.252) / 2}{11{,}790.252 / 22} = 10.69$$

Testing the Stability: Chow Test

$$\sigma_1^2 = \frac{RSS_1}{n_1 - 2} = \frac{1785.032}{10} = 178.5032$$

$$\sigma_2^2 = \frac{RSS_1}{n_2 - 2} = \frac{10,005.22}{12} = 833.7683$$

$$\frac{\hat{\sigma}_1^2 / \sigma_1^2}{\hat{\sigma}_2^2 / \sigma_2^2} \approx F_{[(n_1 - k),(n_2 - k)]}$$

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$$

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{83.7683}{178.5032} = 4.6701$$

Data mining approach

Detecting the presence of unnecessary variables

Suppose we develop a model where we are not sure about X_{ki}

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_{1i}$$

To find that we look at:

1. Test the significance of β_k,
2. Or we if we are not sure about two random variables then we look at F test

However t test and F test should not be used for model building. In the above **approach we assume that we know the model**.

The approach of starting with a smaller model and step by step expansion of that model is called **data mining**

Tests of incorrect functional form and omitted variable

- In practice we never sure that the model adapted is the true model. We look at Signs of estimated coefficients
- Statistical significance
- R squared and adjusted R squared
- F test
- DW statistics

Examination of the residuals

- Examination of the residuals is a good visual diagnostic to detect autocorrelation or heteroscedasiticity and also to detect omitted variable and incorrect functional form

For instance consider cubic cost function

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X^2_{3i} + \beta_3 X^3_{4i} + e_{1i}$$

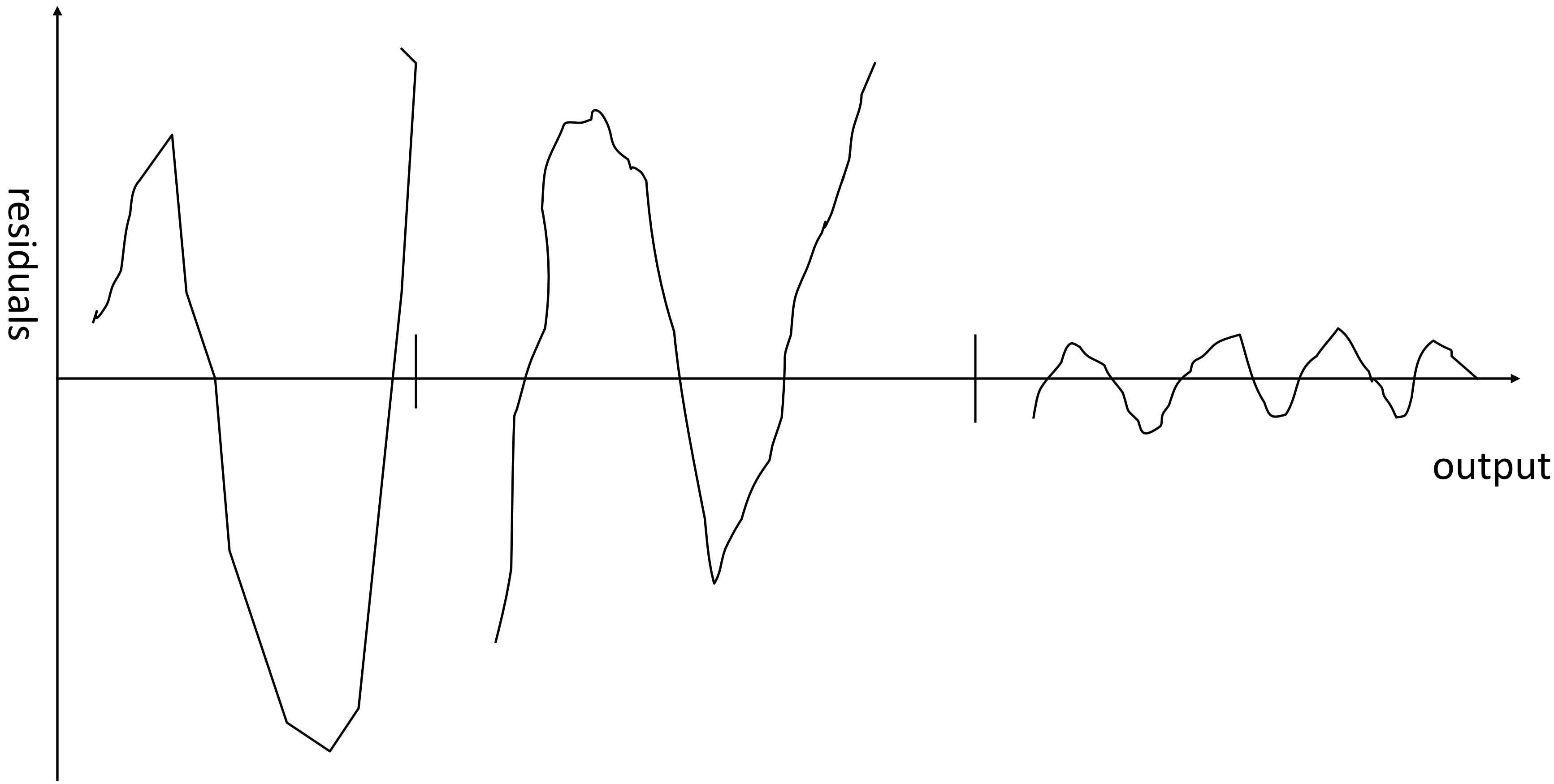
Suppose researcher estimates

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X^2_{3i} + e_{1i}$$

An other researcher estimates

$$Y_i = \beta_1 + \beta_2 X_{2i} + e_{1i}$$

Examination of the residuals



Linear cost function

Quadratic cost function

Cubic cost function

Durbin Watson D statistic

1. From the observed data obtain OLS estimates

2. If it is believed that the model is mis-specified because it excludes a relevant explanatory variable, say Z, order the residuals according increasing values of Z

3. Compute d statistic

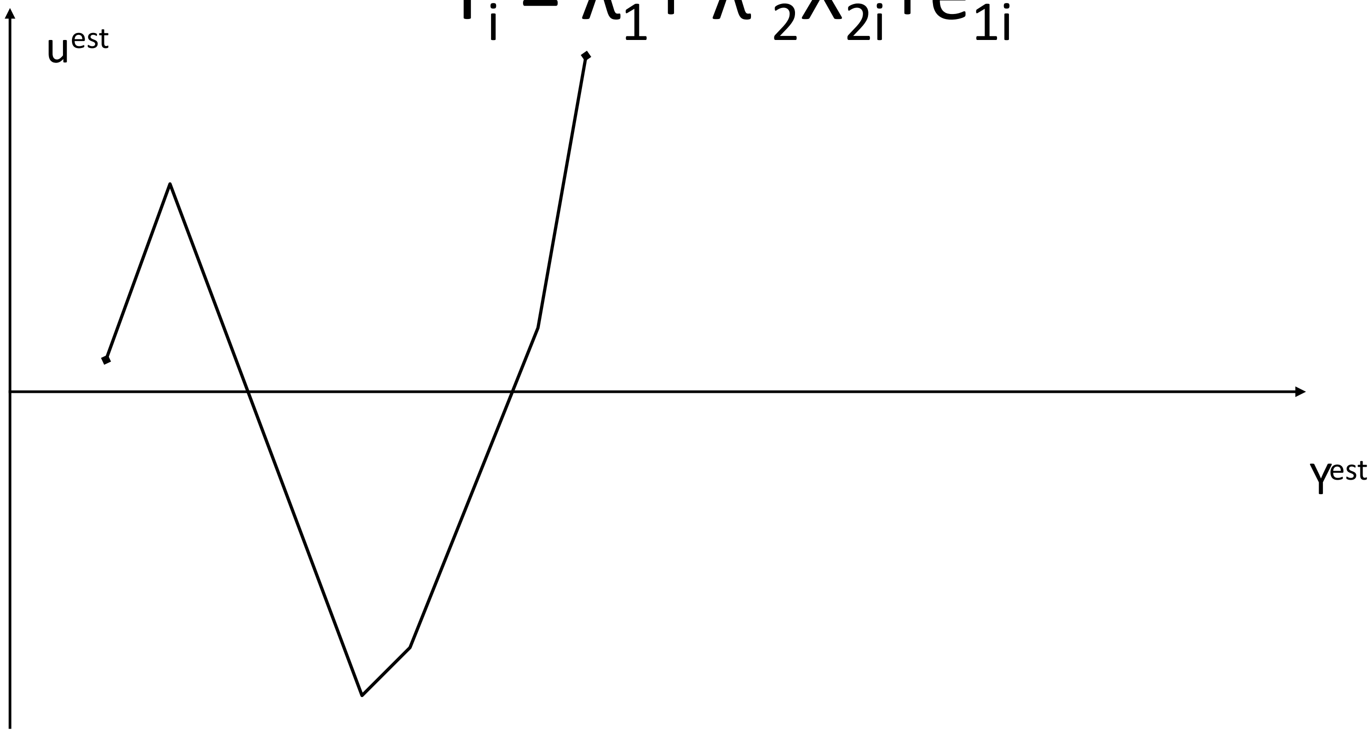
   $d=\sum^{n}_{t=2} (u_t^{est}-u_{t-1}^{est})/ \sum^{n}_{t=1} (u_t^{est})^2$

4. From the Durbin Watson tables find out whether estimated d values is statistically significant, then one can accept the hypothesis of model mis-specification

Ramsey`s RESET
(regression specification error test)

Assume cost function is linear in output where Y total cost and X is the output

$$Y_i = \lambda_1 + \lambda_2 X_{2i} + e_{1i}$$



The idea behind RESET test

From the graph one can see that the mean of error term systematically changes with Y. this suggests that if we introduce estimated Y in the form of the regressor then it should increase R squared. And if on the basis of F test increase in R squared is statistically significant then it would suggest that the model is misperceived.

Steps of Reset test

- From the chosen model obtain estimated Y_i
- Rerun the regression introducing the estimated Y_i thus we run

$$Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 (Y^{est})^2 + \lambda_4 (Y^{est})^3 + e_{1i}$$

- let R^2 obtained from the above model be R^2_{new} and from previous model $R^2_{old.}$ Then we use F test

$$F = \frac{(R^2_{new} - R^2_{old})}{\text{number of new regressors}} : \frac{(1 - R^2_{new})}{(n - \text{number of parameters in new model}}$$

- If the computed F value is significant, say, at 5 percent level, one can accept the hypothesis that the model is mis-specified

LM test for adding the variables

- Estimate the restricted regression by OLS and obtain the residuals, u^{est}
- If in fact the unrestricted regression is the true regression, the residuals obtained from the restricted regression should be related to the squared and cubed output terms, that is X^2_i and X^3_i
- Regress the e^{est} on all regressors

  $e^{est}_{1i} = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X^2 + \lambda_4 X^3 + v$
- For large sample
- $nR^2 \sim$ chi squared distribution with (number of restrictions)
- If the Chi squared value obtained exceeds critical chi squared value at the chosen level of significance we reject the restricted regression

Errors in measurement of dependent variable

$Y^{*}=\alpha+\beta X_i+e_i$

Y^{*} - permanent income

X- current income

$Y^{*}=Y_i+u_i$

$Y^{*}=(\alpha+\beta X_i+e_i)+u_i$

$v=e_i+u_i$

Errors in measurement of dependent variable give unbiased estimates however with larger estimated variances.

Errors in independent variable

$Y_i = \alpha + \beta X^*_i + e_i$

Y_i – current consumption expenditure

X^*_i- permanent income

$X_i = X^*_i + w_i$

$Y^* = (\alpha + \beta X_i + e_i) + w_i$

$v = e_i + \beta w_i$

Errors in measurement of independent variable give both biased estimates and inefficient estimates

Model selection criteria

- R Square: $R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$ $\quad TSS = ESS + RSS$

- Adjusted R square $\bar{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k}$

- AIC $\quad AIC = e^{2k/n}\frac{\sum u_i^2}{n} = e^{2k/n}\frac{RSS}{n}$

- SIC $\quad SIC = n^{k/n}\frac{\sum u_i^2}{n} = n^{k/n}\frac{RSS}{n}$