Model

Selection Criteria

Covered Topics

- Criteria
- Adding/dropping variables
- Double linear against double log
- Linear with different X's

Forecasting Error

Forecasting error = $Y_i - \hat{Y}_i$

Linear model $\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \ldots \hat{\beta}_K X_{Ki}$

Double log

$$\widehat{\ln Y_i} = \hat{\beta}_1 \ln X_{1i} + \hat{\beta}_2 \ln X_{2i} + \ldots \hat{\beta}_K \ln X_{Ki}$$

$$\hat{Y}_i = \exp(\widehat{\ln Y_i})$$

General Selection Criteria (1)

Root Mean Square Error

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

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

General Selection Criteria (2)

Mean Absolute Percentage Error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

Theil's inequality coefficient

$$U = \frac{\sqrt{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}}{\sqrt{\sum_{i=1}^{n} Y_i^2} + \sqrt{\sum_{i=1}^{n} \hat{Y}_i^2}}$$

Relevancy of Variables (1)

Should $Z_1, Z_2, \ldots, Z_M$ be added?

Test Equation

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \delta_1 Z_{1i} + \delta_2 Z_{2i} + \ldots + \delta_M Z_{Mi} + \varepsilon_i$$

Perform an F-test or χ^2-test on

$$H_0 : \delta_1 = \delta_2 = \ldots = \delta_M = 0$$

$$H_1 : \delta_1 \neq \delta 2 \neq \ldots \neq \delta_M \neq 0$$

Relevancy of Variables (2)

Note that $F_{cal} \sim F(M, n-(K+M))$

$$\chi^2_{cal} \sim \chi^2(M)$$

EViews gives the 2-run F-test and the LR test.

Reject H_0 implies that the variables may be relevant. It needs explanation. No harm to the unbiasedness of the estimator of parameters β which have been already included.

Another test equation is the old residuals against all the X's (old & new) w/ a constant term.

Redundancy of Variables (1)

Should $X_{K-M+1}, X_{K-M+2}, \ldots, X_K$ be dropped?

Test Equation

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{K-M} X_{K-M,i} + \beta_{K-M+1} X_{K-M+1,i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

Perform an F-test or χ^2-test on

$$H_0 : \beta_{K-M+1} = \beta_{K-M+2} = \ldots = \beta_K = 0$$

$$H_1 : \beta_{K-M+1} \neq \beta_{K-M+2} \neq \ldots \neq \beta_K \neq 0$$

Redundancy of Variables (2)

Note that $F_{cal} \sim F(M, n-K)$

$\chi^2_{cal} \sim \chi^2(M)$

EViews also gives 2-run F-test and LR test results.

Accept H_0 does <u>not</u> imply that the variables are not relevant. It just says that they are redundant. With the other X's in the model, they are not needed to explain the variation of Y. If they are dropped, there might be estimation bias but no harm to prediction of Y.

MWD Test (1)

MacKinnon-White-Davidson Test

Choose between lin-lin and log-log with same X and Y

<u>Model A</u>

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

<u>Model B</u>

$$\ln Y_i = \gamma_1 \ln X_{1i} + \gamma_2 \ln X_{2i} + \ldots + \gamma_K \ln X_{Ki} + \varepsilon_i$$

MWD Test (2)

Note that R^2 cannot be used to judge.

Given $\hat{Y}_i$ the fitted value from Model A

$\widehat{\ln Y_i}$ the fitted value from Model B

Define

$$Z_{Ai} = \ln \hat{Y}_i - \widehat{\ln Y_i}$$

$$Z_{Bi} = \hat{Y}_i - \exp(\widehat{\ln Y_i})$$

MWD Test (3)

<u>Model A'</u>

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki\,i} + \delta_A Z_{Ai} + \varepsilon_i$$

t-Test on

$H_0 : \delta_A = 0$

$H_1 : \delta_A \neq 0$

Accept H_0 => double lin "encompasses" double log

MWD Test (4)

Model B'

$$\ln Y_i = \gamma_1 \ln X_{1i} + \gamma_2 \ln X_{2i} + \ldots + \gamma_K \ln X_{Ki} + \delta_B Z_{Bi} + +\varepsilon_i$$

t-Test on $H_0 : \delta_B = 0$

$H_1 : \delta_B \neq 0$

Accept H_0 =>double log "encompasses" double line

MWD Test (5)

Conclusion

	$\delta_B = 0$	$\delta_B \neq 0$
$\delta_A = 0$	No clear preference	Double linear is chosen
$\delta_A \neq 0$	Double log is chosen	Neither model good enough

J-test (1)

Davidson-MacKinnon's J-test

Choose between two linear models with different set of explanatory variables but same dependent variable.

R^2 cannot be used either.

Model C $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$

Model D $Y_i = \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \ldots + \gamma_L Z_{Li} + \varepsilon_i$

J-test (2)

Model C'

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki\,i} + \delta_C \hat{Y}_{Di} + \varepsilon_i$$

where $\hat{Y}_i^D$ is the fitted value from Model D

t-Test on $H_0 : \delta_C = 0$

$H_1 : \delta_C \neq 0$

Accept H_0 =>model C "encompasses" model D

J-test (3)

Model D'

$$Y_i = \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \ldots + \gamma_L Z_{Li} + \delta_D \hat{Y}_{Ci} + \varepsilon_i$$

where $\hat{Y}_{Ci}$ is the fitted value from Model C

t-Test on $H_0 : \delta_D = 0$

$H_1 : \delta_D \neq 0$

Accept H_0 =>model D "encompasses" model C

J-test (4)

Conclusion

	$\delta_D = 0$	$\delta_D \neq 0$
$\delta_C = 0$	No clear preference	Model C is chosen
$\delta_C \neq 0$	Model D is chosen	Neither model good enough