Elements of Econometrics. Lecture 7. MLR Model. Predictions.

FCS, 2022-2023

PREDICTION

True model	$Y_i = \beta_1 + \beta_2 X_i + u_i$
Fitted model	$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$
Prediction conditional on X^*	$\hat{Y}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$

$$\text{Prediction Error } PE = Y^* - \hat{Y}^*$$

$$Y^* = \beta_1 + \beta_2 X^* + u^*$$

$$PE = Y^* - \hat{Y}^* = (\beta_1 + \beta_2 X^* + u^*) - \left(\hat{\beta}_1 + \hat{\beta}_2 X^*\right)$$

$$\begin{aligned} \mathrm{E}(PE) &= \mathrm{E}(\beta_1 + \beta_2 X^* + u^*) - \mathrm{E}\left(\hat{\beta}_1 + \hat{\beta}_2 X^*\right) \\ &= \beta_1 + \beta_2 X^* + \mathrm{E}(u^*) - \mathrm{E}\left(\hat{\beta}_1 + \hat{\beta}_2 X^*\right) \\ &= \beta_1 + \beta_2 X^* - \beta_1 - X^* \beta_2 = 0 \end{aligned}$$

Variance of prediction error: proof

$$\sigma_{PE}^2 = \left\{1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right\}\sigma_u^2$$

$$\mathrm{Var}(PE) = \mathrm{Var}(Y^* - \hat{Y}^*) = \mathrm{Var}(Y^*) + \mathrm{Var}(\hat{Y}^*) + 0 = {\sigma_u}^2 + \mathrm{Var}(\hat{\beta}_1 + \hat{\beta}_2 X^*) =$$

$$= \sigma_u^2 + \left(\frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right)\sigma_u^2 = \left(1 + \frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \overline{X})^2}\right)\sigma_u^2$$

$$\mathrm{var}(\hat{\beta}_1 + \hat{\beta}_2 X^*) = \sigma_u^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}\right) + (X^*)^2\frac{\sigma_u^2}{\sum(X_i - \bar{X})^2} + 2X^*\frac{-\overline{X}{\sigma_u}^2}{\sum(X_i - \overline{X})^2} =$$

$$= \left(\frac{1}{n} + \frac{(\bar{X}^2 - 2X^*\overline{X} + (X^*)^2)}{\sum(X_i - \bar{X})^2}\right)\sigma_u^2 = \left(\frac{1}{n} + \frac{(X^* - \overline{X})^2}{\sum(X_i - \bar{X})^2}\right)\sigma_u^2$$

Variance of prediction error: auxiliary proofs

Prove $\mathrm{Cov}(Y^*, \hat{Y}^*) = 0$:

$$\mathrm{Cov}(Y^*, \hat{Y}^*) = \mathrm{Cov}(u*, \hat{\beta}_1 + \hat{\beta}_2 X^*)$$

$$\mathrm{Cov}(u*, \hat{\beta}_2 X^*) = X * \mathrm{Cov}(u*, \beta_2 + \sum_i u_i a_i) = 0 \quad \text{since } u* \text{ and } u_i \text{ are independent}$$

$$\mathrm{Cov}(u*, \hat{\beta}_1) = \mathrm{Cov}(u*, \bar{Y} - \hat{\beta}_2 \overline{X}) = \mathrm{Cov}(u*, \beta_1 + \beta_2 \overline{X} + \overline{u} - \hat{\beta}_2 \overline{X}) = \overline{X}\,\mathrm{Cov}(u*, \hat{\beta}_2) = 0.$$

Prove: $\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \dfrac{-\overline{X}{\sigma_u}^2}{\sum(X_i - \overline{X})^2}$

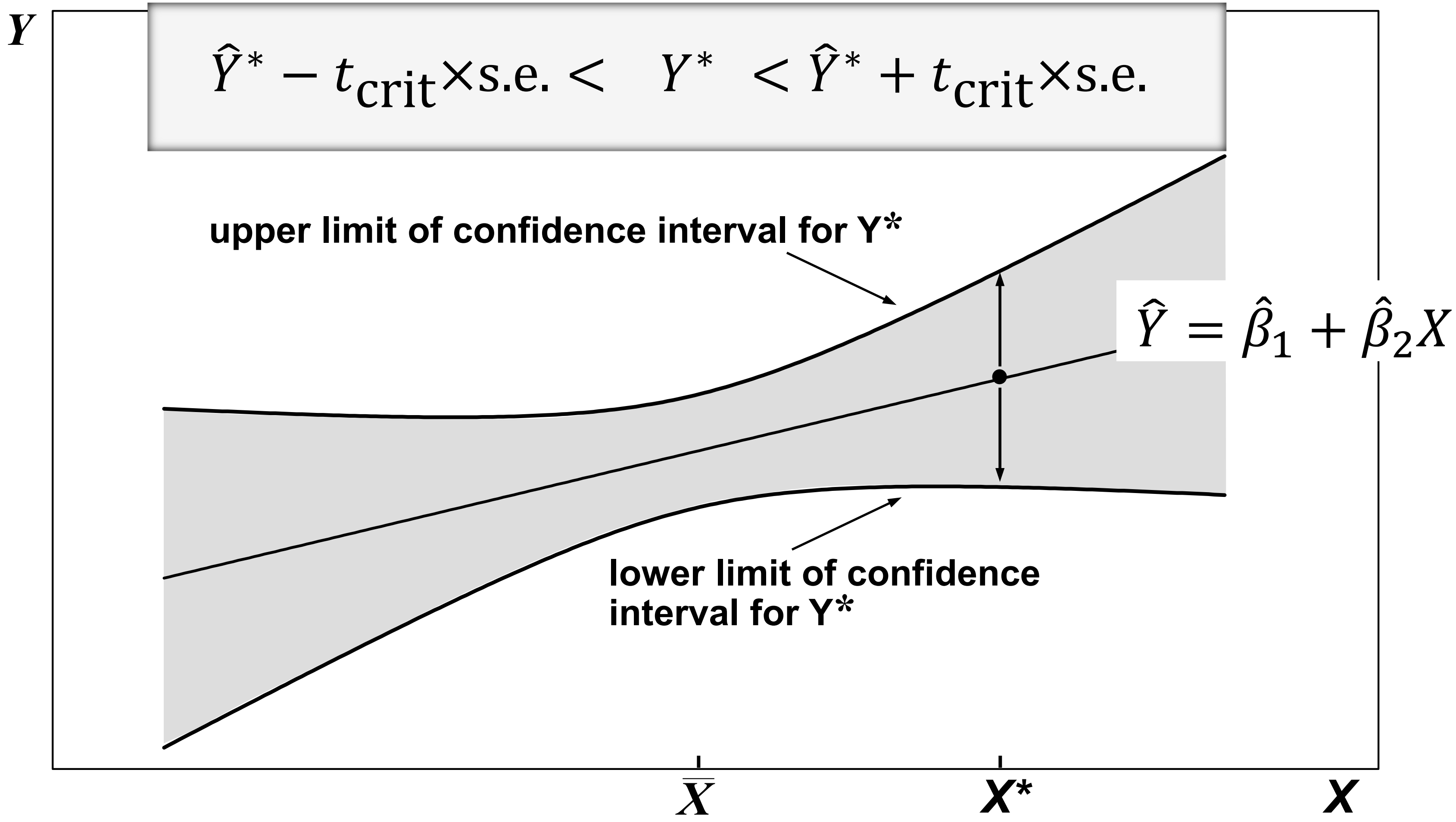
$$\mathrm{Cov}(\overline{Y}, \hat{\beta}_2) = \mathrm{Cov}\left(\sum \frac{1}{n} Y_i, \sum a_j Y_j\right) =$$

$$= \sum \left(\frac{a_i}{n}\right) \mathrm{Var}(Y_i) + \sum_j \left(\frac{a_j}{n}\right) \sum_{i \neq j} \mathrm{Cov}(Y_i, Y_j) = \frac{\sigma_u^2}{n} \sum a_i + 0 = 0$$

$$\mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \mathrm{Cov}(\overline{Y} - \hat{\beta}_2 \overline{X}, \hat{\beta}_2) = \mathrm{Cov}(\overline{Y}, \hat{\beta}_2) - \overline{X}\,\mathrm{Cov}(\hat{\beta}_2, \hat{\beta}_2) =$$

$$= 0 - \overline{X}\,\mathrm{Var}(\hat{\beta}_2) = \frac{-\overline{X}{\sigma_u}^2}{\sum(X_i - \overline{X})^2}$$

PREDICTION



$$\sigma_{PE}^2 = \left\{1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right\}\sigma_u^2 \qquad \text{s.e.}(PE) = \sqrt{\left\{1 + \frac{1}{n} + \frac{(X^* - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right\}s_u^2}$$

PREDICTIONS: GENERAL QUALITY

Root Mean Squared Error

$$\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}$$

Mean Absolute Error

$$\sum_{t=T+1}^{T+h} |\hat{y}_t - y_t| / h$$

. ***Mean Absolute Percent Error***

$$100 \sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{y_t} \right| / h$$

Theil Inequality Coefficient

$$U_1 = \frac{\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / h}}{\sqrt{\sum_{t=T+1}^{T+h} \hat{y}_t{}^2 / h} + \sqrt{\sum_{t=T+1}^{T+h} y_t{}^2 / h}}$$

Theil U_2 Coefficient

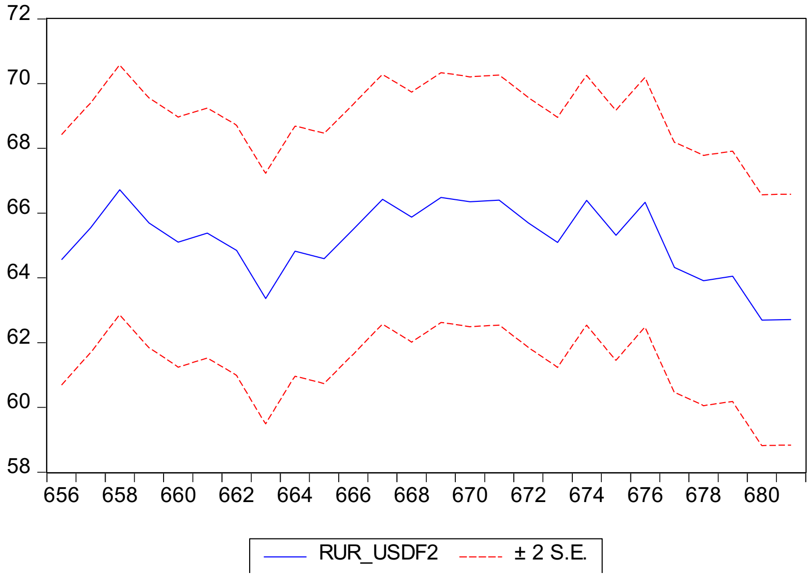
$$U_2 = \sqrt{\frac{\frac{1}{h} \sum (\hat{y}_{T+p} - y_{T+p})^2}{\frac{1}{h} \sum (\Delta y_{T+p})^2}}$$

PREDICTION: RUR/USD Exchange Rate: 12/01/16-31/08/16-06/10/16

Dependent Variable: RUR_USD
Method: Least Squares
Date: 10/05/16 Time: 21:31
Sample: 496 655
Included observations: 160

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	100.3502	0.995850	100.7683	0.0000
OIL_BRENT(-2)	-0.739957	0.022991	-32.18508	0.0000

R-squared	0.867658	Mean dependent var	68.67378



Forecast: RUR_USDF2	
Actual: RUR_USD	
Forecast sample: 656 681	
Included observations: 26	
Root Mean Squared Error	1.217824
Mean Absolute Error	1.037165
Mean Abs. Percent Error	1.615266
Theil Inequality Coefficient	0.009408
Bias Proportion	0.531535
Variance Proportion	0.032971
Covariance Proportion	0.435494
Theil U2 Coefficient	2.685163
Symmetric MAPE	1.598604

Forecast for 06.10.16:

62.7

Official ER 06.10.16:

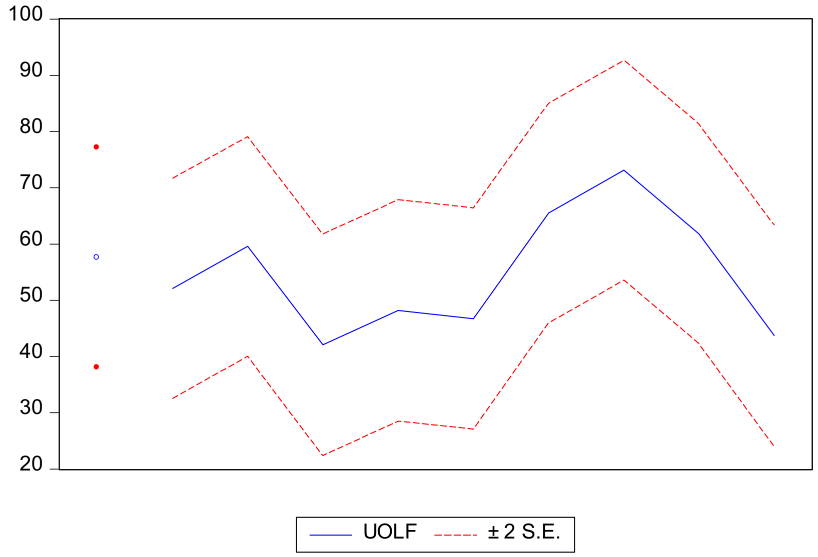
62.46

PREDICTION: ICEF STUDENTS UoL GRADES, Elements of Econometrics, 2019

Dependent Variable: UOL
Method: Least Squares
Date: 10/05/19 Time: 21:12
Sample (adjusted): 1 199
Included observations: 149 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	37.84472	1.768514	21.39916	0.0000
SEM1	0.224676	0.071463	3.143929	0.0020
MARCH	0.499192	0.080407	6.208298	0.0000

R-squared	0.641020	Mean dependent var	62.94631
Adjusted R-squared	0.636102	S.D. dependent var	16.12401
S.E. of regression	9.726637	Akaike info criterion	7.407542
Sum squared resid	13812.69	Schwarz criterion	7.468024
Log likelihood	-548.8619	Hannan-Quinn criter.	7.432115
F-statistic	130.3538	Durbin-Watson stat	2.012132
Prob(F-statistic)	0.000000		



Forecast: UOLF	
Actual: UOL	
Forecast sample: 201 211	
Adjusted sample: 201 211	
Included observations: 10	
Root Mean Squared Error	9.051834
Mean Absolute Error	7.938863
Mean Abs. Percent Error	14.03293
Theil Inequality Coef.	0.077633
Bias Proportion	0.064915
Variance Proportion	0.003274
Covariance Proportion	0.931811
Theil U2 Coefficient	0.566297
Symmetric MAPE	14.33669

PREDICTION with logarithmic functions

- **Predicting *Y* when log(*Y*) is the dependent variable**

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

$$\Rightarrow \quad y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) \exp(u)$$

- Under the additional assumption that *u* is independent of $X_1,\ldots,X_k$:

$$\Rightarrow \quad E(y|\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) E(\exp(u))$$

$$\Rightarrow \quad \hat{y} = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k)(\frac{1}{n}\sum_{i=1}^{n} \exp(\hat{u}_i))$$