Predictions

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

$$\hat{y}_{n+1} = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_{n+1}$$

$$E(\hat{y}_{n+1}) = \beta_1 + \beta_2 \cdot x_{n+1}$$

$$E(y_{n+1}) = E(\beta_1 + \beta_2 \cdot x_{n+1} + \varepsilon_{n+1}) =$$

$$\beta_1 + \beta_2 \cdot x_{n+1}$$

$$E(\hat{y}_{n+1} - y_{n+1})^2 = E(\hat{\beta}_1 + \hat{\beta}_2 x_{n+1} -$$

$$\beta_1 - \beta_2 x_{n+1} - \varepsilon_{n+1})^2 =$$

$$E((\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2) \cdot x_{n+1} - \varepsilon_{n+1})^2.$$

$$= E(\hat{\beta}_1 - \beta_1)^2 + x_{n+1}^2 E(\hat{\beta}_2 - \beta_2)^2 + E(\varepsilon_{n+1})^2$$

$$2x_{n+1} E((\hat{\beta}_1 - \beta_1) \cdot (\hat{\beta}_2 - \beta_2)) -$$

$$\varepsilon \sim N(0, \sigma_\varepsilon^2)$$

$$-2x_{n+1} E((\hat{\beta}_2 - \beta_2) \cdot \varepsilon_{n+1}) -$$

$$-2E((\hat{\beta}_1 - \beta_1)\varepsilon_{n+1}) =$$

$$Var(\hat{\beta}_1) + x_{n+1}^2 Var(\hat{\beta}_2) + \sigma_\varepsilon^2 -$$

$$2x_{n+1} Cov(\hat{\beta}_1, \hat{\beta}_2) - 0 - 0 \; =$$

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{bmatrix} \dfrac{\sigma^2 \Sigma X_i^2}{n \Sigma (X_i - \bar{X})^2} & \dfrac{-\bar{X}\sigma^2}{\Sigma (X_i - \bar{X})^2} \\ \dfrac{-\bar{X}\sigma^2}{\Sigma (X_i - \bar{X})^2} & \dfrac{\sigma^2}{\Sigma (X_i - \bar{X})^2} \end{bmatrix}$$

$$= \frac{\sigma^2 \Sigma X_i^2}{n \Sigma (X_i - \bar{X})^2} + x_{n+1}^2 \frac{\sigma^2}{\Sigma (X_i - \bar{X})^2}$$

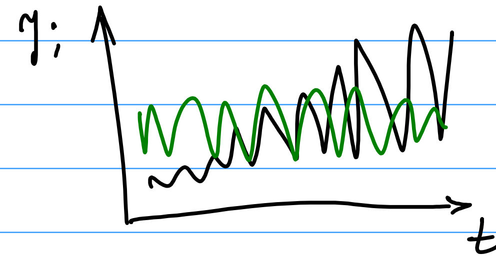
$$+ 2x_{n+1} \cdot \frac{-\bar{X}\sigma^2}{\Sigma (X_i - \bar{X})^2} =$$

$$\sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{X})^2}{\Sigma (x_i - \bar{X})^2} \right) = \delta^2$$

$\delta = se(\hat{\varepsilon}_{n+1})$ s.e. of prediction error

$$y_{n+1} : \{ \hat{y}_{n+1} \pm t_{n-2, \alpha/2} \, se(\hat{\varepsilon}_{n+1}) \}$$

Log transformations



Linear: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

$\frac{dy}{dx} = \beta_2 \Rightarrow x \uparrow 1 \quad y \uparrow \beta_2$

Log-log: $\ln y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$ $\quad y_i = \beta_1^* \cdot x^{\beta} \cdot \varepsilon_i$

$\frac{d \ln y_i}{d \ln x_i} = \frac{100 \cdot dy_i / y_i}{100 \; dx_i / x_i} = \beta_2$

$\left\{ \begin{array}{l} Y = A \cdot K^{\alpha} L^{1-\alpha} \\ H_0: \beta_2 + \beta_3 = 1 \end{array} \right.$

$x \uparrow 1\% \quad y_i \uparrow \beta\%$

Lin-log: $y_i = \beta_1 + \beta_2 \ln x_i + \varepsilon_i$

$\frac{dy_i}{d \ln x_i} = \beta_2 \qquad \frac{dy_i}{100 \; dx_i / x_i} = \beta_2 / 100 \qquad x \uparrow 1\% \; y \uparrow \frac{\beta_2}{100}$

Log-lin: $\ln y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$

$\frac{d \ln y_i}{dx_i} = \beta_2 \qquad \frac{100 \cdot dy_i / y_i}{dx_i} = 100 \cdot \beta_2 \qquad x \uparrow 1 \quad y_i \uparrow 100 \beta_2 \%$

Problem 7. (UoL Exam) A regression of consumption (C) on income (Y) and unemployment (U) (all variables are index numbers) using annual data 1961-82 for the UK produced the following results:

$$\hat{C}_t = \underset{(2817.0)}{17880} + \underset{(0.026)}{0.7527} Y_t + \underset{(0.798)}{0.930} U_t, \quad R^2 = 0.992 \qquad (1)$$

(figures in brackets are standard errors) with a table of correlation coefficients between variables of:

	C	Y	U
C	1.00	0.996	0.783
Y	0.996	1.00	0.771
U	0.783	0.771	1.00

log log %

α β_1 $\hat{\beta}_2$

$n = 22$

a) interpret and test $\quad t_{19}^{2.5\%} = 2.09$

$t_{\hat{\alpha}} = \frac{17\ldots}{2\ldots} = 6{,}35 \quad 1\%$

$t_{\hat{\beta}_1} = \frac{0{,}75}{0.026} = 28{,}95 \quad 1\%$

$t_{\hat{\beta}_2} = \frac{0{,}93}{0{,}8} = 1{,}16 \quad 5\%$

b) $\nearrow$ C | Y assuming $\beta_2 < 0 \Rightarrow$ lower $\hat{\beta}_1$

c) assume: $\beta_1 \geq 0$, $\beta_2 \leq 0$

$H_0: (\beta_1 \geq 0)\ \beta_1 = 0 \qquad H_0: \beta_1 = \beta_2 = 0$

$H_a: \beta_1 < 0 \qquad \hookrightarrow$ F-test

Quadratic and Interactive terms

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$$

$$\frac{d y_i}{d x_i} = \beta_2 + 2\beta_3 \cdot x_i$$

β_2 shows rate of change of y if $x_i = 0$

$$y_i = \beta_1 + (\beta_2 + \beta_3 x_i) x_i + \varepsilon_i$$

β_3 shows rate of change of x_i

per unit change of x_i

Problem 1. (UoL Exam).

The rise in prices for public transport leads to lower corporate earnings, as people tend to choose cheaper alternatives. The student tries to find the best form of dependence of the volume of transportation T_i of some 50 transportation companies (in millions of dollars) from the prices of transportation P_i (in cents per one kilometer of transportation). She runs regressions (1-4) (linear, logarithmic and semi-logarithmic functions), she also runs two auxiliary regressions (5-6) performing Zarembka transformation (variable TZ_i is defined as $TZ_i = T_i / \sqrt[n]{T_1 \cdot T_2 \cdot \ldots \cdot T_n}$):

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	T_i	T_i	$\log(T_i)$	$\log(T_i)$	TZ_i	TZ_i
Independent variable\Constant	8.74	12.26	2.175	2.635	1.171	1.641
P_i	-0.339	-	-0.0045		-0.0045	
$\log(P_i)$	-	-1.362	-	-0.179	-	-0.179
R^2	0.638	0.738	0.665	0.755	0.638	0.738
RSS	4.481	3.247	0.068	0.051	0.080	0.058

Question 2. (ICEF Exam)

An employee of a real estate agency in a Russian city with a developed subway network is interested in estimating of the influence of the distance from the city center $CENTER_i$ (in kilometers) on the price of an two-room apartment in millions of rubles. Based on the data of 21 apartments sold during a period under consideration she runs a regression.

$$\underset{(0.88)}{\widehat{PRICE}_i = 12.39} - \underset{(0.10)}{0.20} \cdot CENTER_i \qquad R^2 = 0.17 \quad RSS = 103.4 \qquad \textbf{(1)}$$

The realtor, not satisfied with the obtained result, decided to take into account the additional factor – the distance to the nearest subway station $METRO_i$ (also in kilometers).

$$\underset{(0.97)}{\widehat{PRICE}_i = 13.71} - \underset{(0.09)}{0.22} \cdot CENTER_i - \underset{(0.25)}{0.58} \cdot METRO_i \qquad R^2 = 0.37 \quad RSS = 79.29 \qquad \textbf{(2)}$$

During the discussion at the workshop, the realtor received advice from a colleague to use Ramsey's test for this equation. Since the realtor was not experienced enough in econometrics, a colleague helped her calculate appropriate equation (using in the right side of (3) estimated values $\widehat{PRICE}_i^*$ from equation (2):

$$\underset{(6.04)}{\widehat{PRICE}_i = 0.023} + \underset{(0.18)}{0.13} \cdot CENTER_i + \underset{(0.47)}{0.35} \cdot METRO_i + \underset{(0.033)}{0.07} \cdot (\widehat{PRICE}_i^*)^2 \qquad R^2 = 0.51 \quad RSS = 60.64 \qquad \textbf{(3)}$$

Then the colleague helped her to estimate a new equation

$$\underset{(0.10)}{\log \widehat{PRICE}_i = 2.62} - \underset{(0.0095)}{0.019} \cdot CENTER_i - \underset{(0.026)}{0.059} \cdot METRO_i \qquad R^2 = 0.32 \quad RSS = 0.8448 \qquad \textbf{(4)}$$

and did Ramsey's test again (using in the right side of (5) estimated values $\log \widehat{PRICE}_i^{**}$ from equation (4):

$$\underset{(1.53)}{\log \widehat{PRICE}_i = 0.62} + \underset{(0.039)}{0.030} \cdot CENTER_i + \underset{(0.11)}{0.084} \cdot METRO_i + \underset{(0.0088)}{0.012} \cdot (\log \widehat{PRICE}_i^{**})^2 \qquad R^2 = 0.39 \quad RSS = 0.7672 \qquad \textbf{(5)}$$

She estimated non-linear regression (4) using logarithm of dependent variable

$$\underset{(0.10)}{\log \widehat{PRICE}_i = 2.62} - \underset{(0.0095)}{0.019} \cdot CENTER_i - \underset{(0.026)}{0.059} \cdot METRO_i \qquad R^2 = 0.32 \quad RSS = 0.8448 \qquad (4)$$

and evaluates Ramsey test again

$$\underset{(1.53)}{\log \widehat{PRICE}_i = 0.62} + \underset{(0.039)}{0.030} \cdot CENTER_i + \underset{(0.11)}{0.084} \cdot METRO_i + \underset{(0.0088)}{0.012} \cdot (\log \widehat{PRICE}_i^{**})^2 \qquad R^2 = 0.39 \quad RSS = 0.7672 \qquad (5)$$