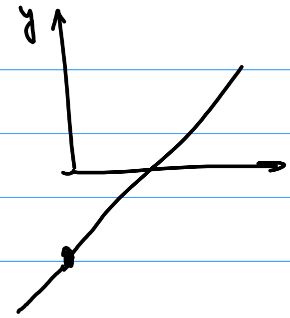
Transformation of Variables

(1) $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$



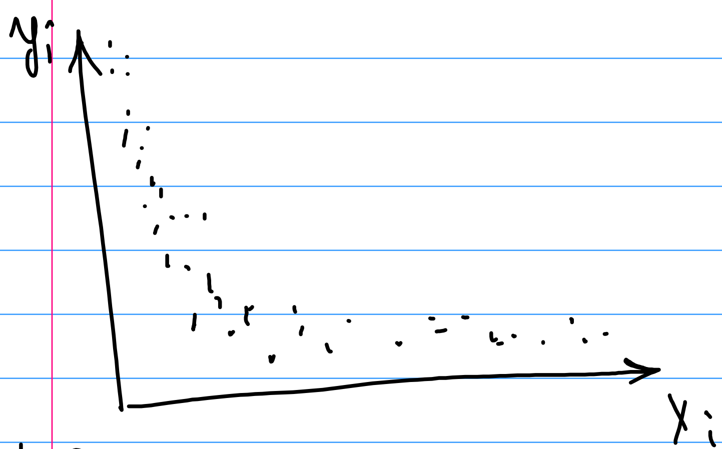
$\beta_0 : \quad E(y_i \mid X_i = 0) = \beta_0$

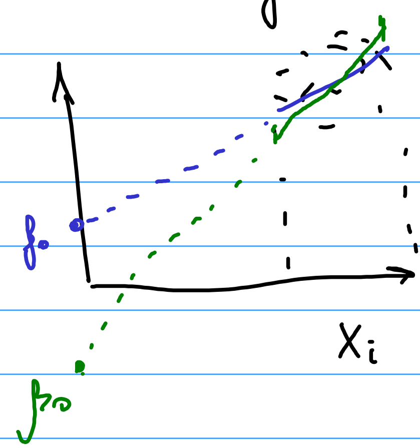
$\beta_1 : \quad X \uparrow 1 \qquad y \uparrow \beta_1$

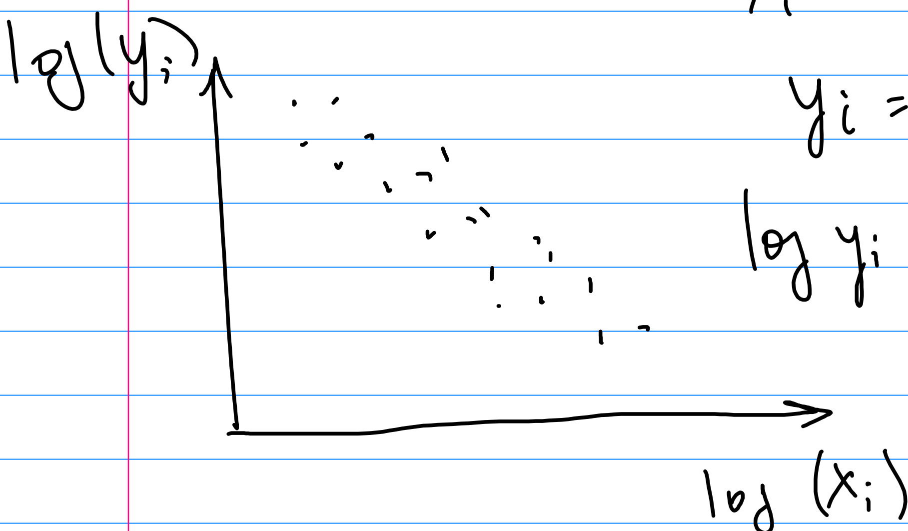
"other being fixed"

$\hat{\beta}_1$: "On average"

(2)







$y_i = \beta_0 \cdot X^{\beta_1} \cdot \varepsilon_i$

$$\log y_i = \underbrace{\log \beta_0}_{=\beta_0^*} + \beta_1 \log x_i + \underbrace{\log \varepsilon_i}_{=u_i}$$

$\log \varepsilon = u_i \sim N(0, \sigma_u^2)$

$\varepsilon_i \sim LN(0, \sigma_u^2)$

1) $y_i = \beta_0 x^{\beta_1} \cdot \varepsilon_i$ (mult)

2) $y_i = \beta_0 x^{\beta_1} + \varepsilon_i$ (add)

$$\beta_1 : \quad d \log y_i \Big/ d \log x_i = 100 \cdot \frac{d y_i}{y_i} \% \Big/ 100 \frac{d x_i}{x_i} \% \approx E_{y|x} = \beta_1$$

$$x \uparrow \% \Rightarrow y \uparrow \beta_1 \%$$

(3) $y_i = \beta_1 + \beta_2 \log x_i + \varepsilon_i$

$$d y_i \Big/ d \log x_i = d y_i \Big/ 100 \frac{d x_i}{x_i} \% = \frac{\beta_2}{100}$$

$$x \uparrow 1\% \Rightarrow y \uparrow \beta_2 / 100$$

(4) $\log y_i = \beta_1 + \beta_2 x_i + \varepsilon$

$$100 \frac{d y_i}{y_i} \% \Big/ d x_i = \beta \cdot 100$$

$$x \uparrow 1 \quad y \uparrow \beta \cdot 100 \%$$

Quadratic term

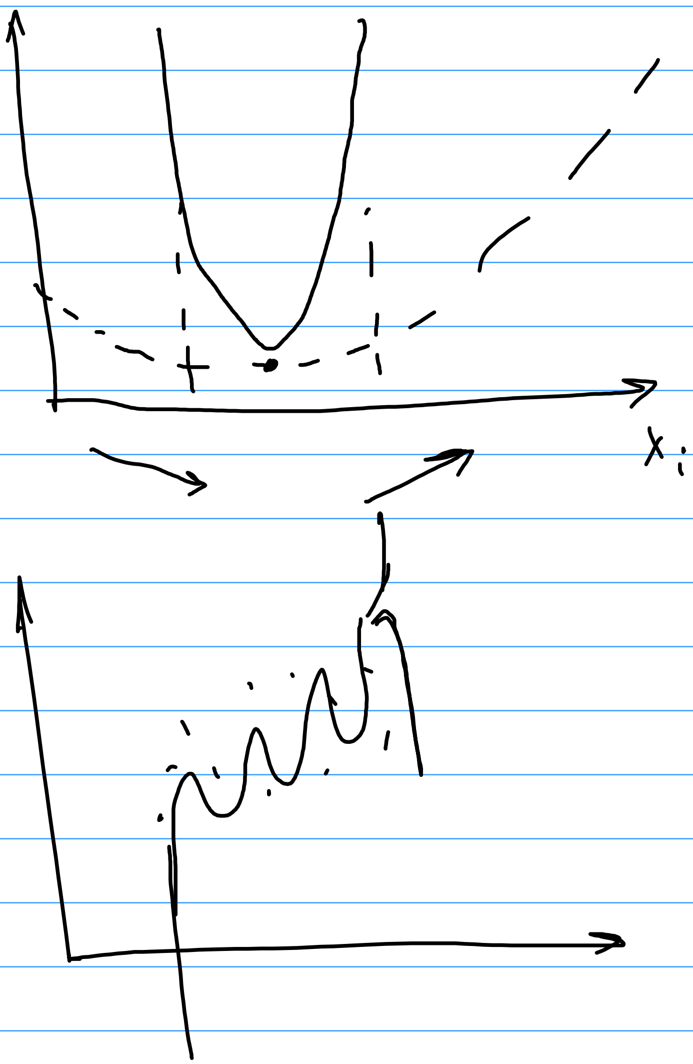
$$y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$$

$$dy_i / dx_i = \beta_1 + \beta_2 \cdot 2x_i$$

β_1 : $x_i \uparrow$ $y_i \uparrow_{\beta_2}$ if $x_i = 0$

β_2 : sign – direction of effect

value – effect size



Multiplicative term

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + \epsilon_i$$

$$dy_i / dx_{1i} = \beta_1 + \beta_3 \cdot x_{2i}$$

β_1 : $x_{1i} \uparrow 1$ $\quad y_i \uparrow \beta_1$ $\quad x_{2i} = 0$

β_2 : $x_{2i} \uparrow 1$ $\quad y_i \uparrow \beta_2$ $\quad x_{1i} = 0$

$$x_i^* = x_i - \bar{x}$$

$$y_i^* = y_i - \bar{y}$$

β_2 : $x_{2i} \uparrow 1$ $\quad y_i \uparrow \beta_2$ $\quad X_{1i} = \bar{X}$

$$\begin{cases} x_{1i}^* = 0 \\ x_i - \bar{x} = 0 \end{cases}$$

RESET Test

$$\hat{y} \Leftarrow y_i^R = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i$$

$\hat{y}^2$

$\leq$?

$(X_1 + \dots + X_k)^2$

$$y^{UR} = y^R + \sum \gamma_j x_{ji}^2 + \sum \theta x_i x_j$$

F test for linear restrictions

$$y = y^R + \gamma_2 \hat{y}_i^2 + \dots + \gamma_p \hat{y}_i^p$$

F test for $H_0: \gamma_2 = \dots = \gamma_p = 0$

Box-Cox (Zarembka)

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 $TZ_i = T_i / \sqrt[n]{T_1 \cdot T_2 \cdot \ldots \cdot T_n}$):

$\frac{dy}{dx}$

?

	(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

(a) Explain the differences in the values of a slope coefficient in regression (1) and (4) giving interpretation to both regressions.

(b) Explain the differences in the values of a slope coefficient in regression (2) and (3) giving interpretation to both regressions.

(c) Explain using some math why your interpretation of regression (4) is correct using different methods. Do the same for regressions 2-3.

(d) Which pairs of regression are comparable directly without Zarembka transformation). Which regressions becomes comparable after Zarembka transformation? Compare some regressions performing appropriate tests.

Box - Cox: $\frac{n}{2}\left|\log\frac{RSS_1}{RSS_2}\right| \sim \chi^2_1$

$\frac{50}{2}\left|\log\frac{0{,}06}{0{,}05}\right| \approx 3{,}215$

Box - Cox: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$

$F(\cdot) = \frac{y^\lambda - 1}{\lambda}$ $\begin{cases} \lambda = 1 & [y-1] \longrightarrow \text{lin} \\ \lambda = 0 & \left[\frac{0}{0}\right] \to \log y \to \log \end{cases}$

$F(y) = y^* = \frac{y^{\lambda_1} - 1}{\lambda_1}$ $F(x) = x^* = \frac{x^{\lambda_2} - 1}{\lambda_2}$

$\hat{\lambda}_1, \hat{\lambda}_2$ $y^* = \beta_1 + \beta_2 x^* + \varepsilon_i$

$\lambda_1 = \lambda_2 = 1$ lin

$\lambda_1 = \lambda_2 = 0$ log

$\lambda_1 = 1$ $\lambda_2 = 0$ lin-log

$\lambda_1 = 0$ $\lambda_2 = 1$ log-lin