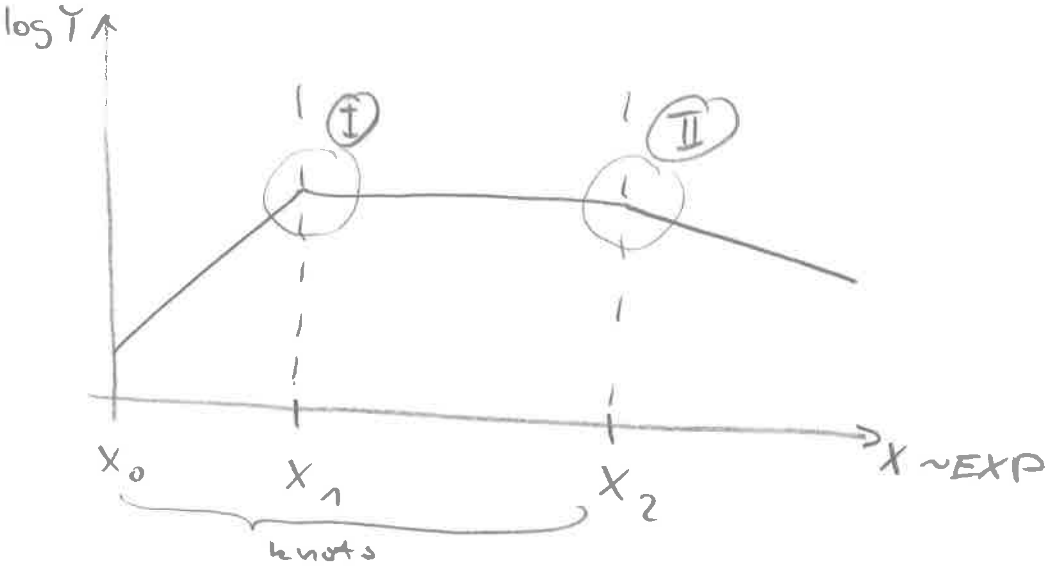
SPLINES

- use if theory does not predict fc. form
- use if continuity f(x) needed

• CASE of EXPerience ~ EXP profile



Choice of knots:

- arbitrary
- other rules

$\bar{X}_0, \bar{X}_1, \bar{X}_2 \sim$ knots $=$ const

$D_1 = 1$ if $\bar{X}_0 \leq X < \bar{X}_1$

$D_2 = 1$ if $\bar{X}_1 \leq X < \bar{X}_2$

$D_3 = 1$ if $\bar{X}_2 < X$

$$Y = [a_1 + b_1(X - \bar{X}_0)]D_1 + [a_2 + b_2(X - \bar{X}_1)]D_2 + [a_3 + b_3(X - \bar{X}_2)]D_3 + u$$

Impose continuity at I & II

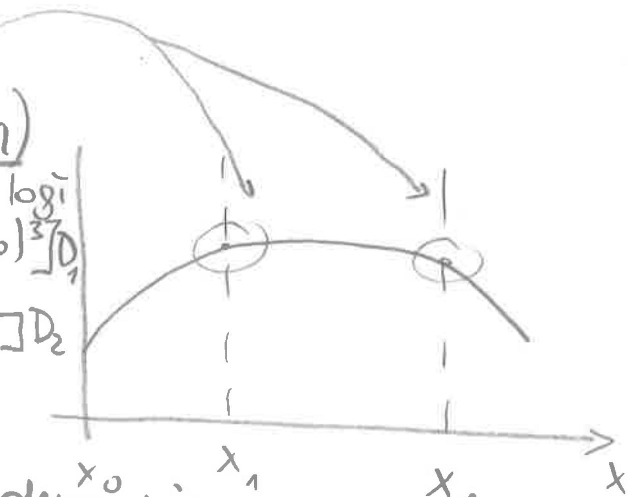
$$a_2 = a_1 + b_1(\bar{X}_1 - \bar{X}_0)$$

$$a_3 = a_2 + b_2(\bar{X}_2 - \bar{X}_1)$$

• POLYNOMIAL SPLINE (smooth)

$$Y = [a_1 + b_1(X - X_0) + c_1(X - X_0)^2 + d_1(X - X_0)^3]D_1 + [a_2 \ldots \ldots]D_2$$

$\vdots$



Impose continuity & continuity of 1st & 2nd derivatives

$\to a_2 = a_1 + b_1(x_1 - x_0) + c_1(x_1 - x_0)^2 + d_1(x_1 - x_0)^3$

$\to b_2 = \ldots$

simple formula (for general specification)

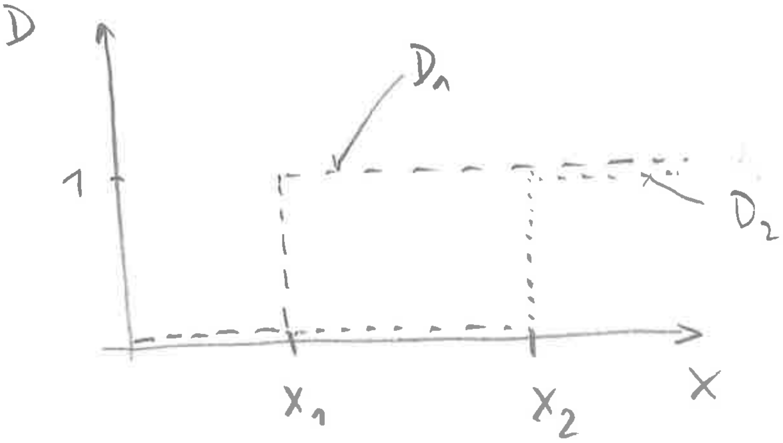
$$\boxed{Y = a_1 + b_1(X - X_0) + c_1(X - X_0)^2 + d_1(X - X_0)^3 + \sum_{i=1}^{k} (d_{i+1} - d_i)(X - X_i)^3 D_i^*}$$

$$\boxed{\begin{aligned} D_i^* &= 1 \quad \text{if} \quad X \geq X_i \\ &= 0 \quad \text{otherwise} \end{aligned}}$$

marginal definition

$k+1 = 3 \sim$ # intervals

$$\underbrace{a_1, b_1, c_1, d_1, d_2, \ldots d_k}_{\text{parameters}}$$



- more X's can be defined by splines

CASE of 1st order:

$D_1 = 1(X \geq X_1)$ $\quad$ $D_2 = 1(X \geq X_2)$

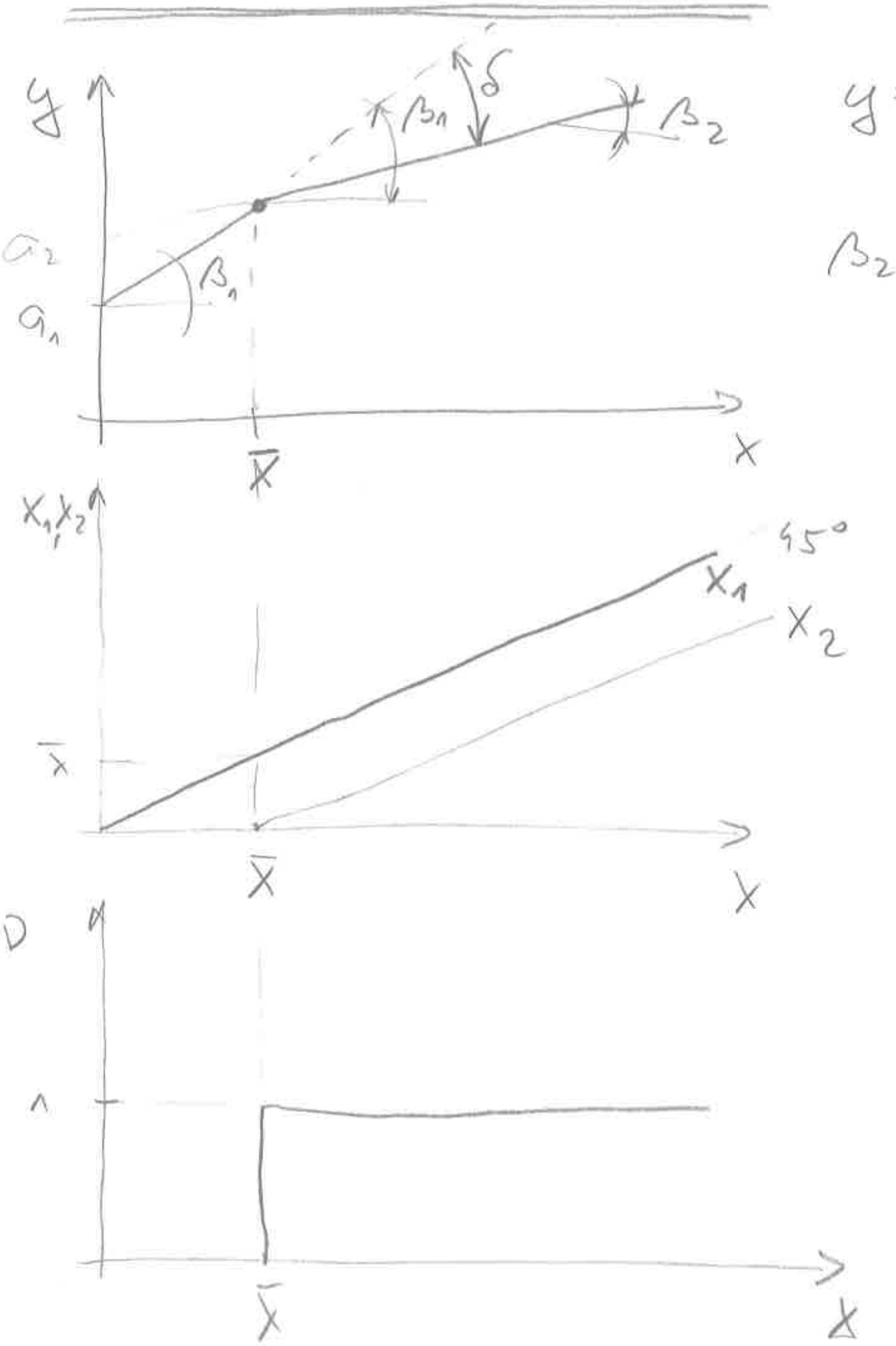
$$Y = a + b_1(X - X_0) + b_2(X - X_1)D_1 + b_3(X - X_2)D_2$$

CASE of 2nd order:

$$Y = a + b_1(X - X_0) + c_1(X - X_0)^2 + c_2(X - X_1)^2 D_1 + c_3(X - X_2)^2 D_2$$

CASE of 3rd order:

tuition on SPLINES



$$y = a_1 + \beta_1 \underset{\downarrow \atop X_1}{X} + \underbrace{(X-\bar{X})\,D(x>\bar{x})}_{X_2}\,\delta$$

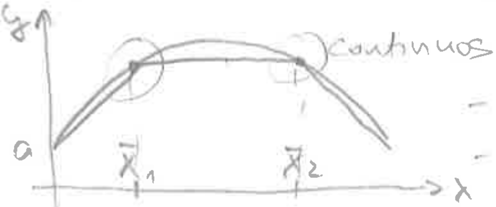
$$\beta_2 = \beta_1 + \delta$$

or

$$\cancel{y = a_1 + \beta_1 X\, D(X \le \bar{X}) +}$$
$$\cancel{+\, a_2 + (X-\bar{X})\, D(X > \bar{X})\, \beta_2}$$

Notes on estimation

Splines → $y = a + b_1 x + b_2(x - \bar{x}_1) D[x > \bar{x}_1] + b_3(x - \bar{x}_2) D[x > \bar{x}_2]$



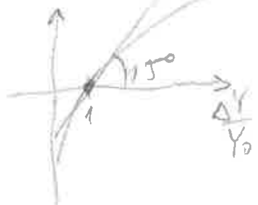
D ... indicator fcion = dummy

- polynomial versions possible
- possible for more x's
- regions of x by definition or optional

$\ln y = \beta x + \delta D \; : \; D = 1, 0$

$$\left.\begin{array}{l} \ln y_0 = \beta x + 0 \\ \ln y_1 = \beta x + \delta \end{array}\right\} \ln y_1 - \ln y_0 = \ln \frac{y_1}{y_0} = \ln\left(\frac{Y_0 + \Delta Y}{Y_0}\right) = \ln\left(1 + \frac{\Delta Y}{Y_0}\right) = \delta$$

$$\% \Delta Y \equiv \frac{\Delta Y}{Y_0} = e^{\delta} - 1 \implies \delta \sim \% \text{ for small } \delta$$



Ability bias

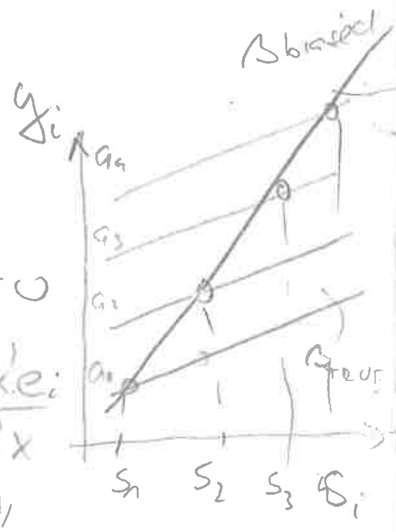
$$y_i = \beta x_i + \underbrace{a_i + e_i}_{\varepsilon_i} \; , \quad \mathrm{cov}(a_i, x_i) > 0$$

$$\mathrm{cov}(a_i, e_i) = \mathrm{cov}(x_i, e_i) = 0$$

$$\hat{\beta}_{OLS} = \frac{x'y}{x'x} = \frac{x'(\beta x + a + e)}{x'x} = \beta + \frac{x'a}{x'x} + \frac{x'e_i}{x'x}$$

$\frac{x'a}{x'x} \to > 0 \to$ bias

$\frac{x'e_i}{x'x} \to 0$



- error in variables

RHS : EDU < recall error, imputation error, drop outs } ⇒ bias ↗ ↘

LHS : Y ... recall → no bias

taxes → bias

missing

$$EXP = AGE - EDU - 6 - h$$

- specification EDU < years ? levels

Errors in variables

$$y = \beta x + \varepsilon \quad \text{true model}$$

$$x^* = x + e$$

$$\hat{\beta}_{OLS} = (x^{*\prime}x^*)^{-1}x^{*\prime}y$$

$$= (x^{*\prime}x^*)^{-1}[x+e]'[\beta x+\varepsilon]$$

$$= \underline{\underline{\frac{\beta\sigma_x^2 + \cancel{x'\varepsilon} + \cancel{\beta e'x} + \cancel{e'\varepsilon}}{\sigma_x^2+\sigma_e^2}}}$$

$x \perp e$, $e \perp \varepsilon$

$\varepsilon \perp x$

$e \sim$ meas. error

$x^* \sim$ actual x

$x \sim$ mismeasured x

$$x^{*\prime}x^* = (x+e)'(x+e)$$

$$= \underline{\underline{\sigma_x^2 + \sigma_e^2}}$$

$$x'x = x^{*\prime}(x^*+e)$$

$$\cancel{= \underline{\underline{\sigma_x^2}}}$$