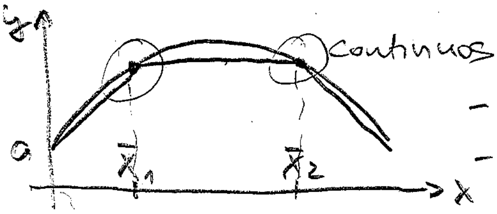
Notes on estimation

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 fcio = dummy

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

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

$$\left.\begin{aligned} \ln y_0 &= \beta x + 0 \\ \ln y_1 &= \beta x + \delta \end{aligned}\right\} \ln y_1 - \ln y_0 = \ln\frac{y_1}{y_0} = \ln\left(\frac{y_0+\Delta y}{y_0}\right) = \ln\left(1+\frac{\Delta y}{y_0}\right) = \delta$$

$$\frac{\Delta y}{y_0} = e^{\delta} - 1 \Rightarrow \delta \sim \% \text{ for small } \delta$$

ability bias

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

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

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

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

$$EXP = AGE - EDU - 6 \; (-h) \text{ missing}$$

- specification EDU < years, levels ?

Errors in variables

$$y = \underbrace{\beta x^* + \varepsilon}$$

$$x = \underbrace{x^* + e}$$

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

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

$$= \frac{\beta \sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2}$$

$e \sim$ meas. error

$x^* \sim$ actual x

$x \sim$ mismeasured x

$$x'x = (x^* + e)(x^* + e)$$

$$= \sigma_{x^*}^2 + \sigma_e^2$$

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

$$= \sigma_{x^*}^2$$

NSTRUMENTAL VARIABLES

- if $cov(X, \varepsilon) \neq 0$ → bias, inconsistency
- when:
  - measurement error in X
  - omitted variable (ability): $cov(A, \varepsilon) \neq 0$
    $cov(A, EDU) \neq 0$
  - autoregression
  - endogenous by principle (econ)
  - lagged var as RHS.
- Solution → find instrument Z

$$cov(Z, X) \neq 0$$
$$cov(Z, \varepsilon) = 0$$

- Procedure 2SLS (P. Kennedy pp. 174-175)

1 stage: regress (OLS) $X_i = \gamma Z_i + \nu_i$

predict $\hat{X}_i = \hat{\gamma} Z_i$

2nd stage: regress (OLS) $Y_i = \hat{X}_i \beta + \varepsilon$ not $\cancel{Y_i = Z_i \beta + \varepsilon}$

$\dots S : cov(X, \varepsilon) = 0$

$$Y = X\beta + \varepsilon \qquad / \cdot X'$$

$$X'Y = X'X\hat{\beta} + X'\varepsilon \to 0 \qquad / (X'X)^{-1}$$

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'y$$

$Y =$

$OLS : cov(X, \varepsilon) \neq 0$

$$Y = X\beta + \varepsilon$$

$cov(X, \varepsilon) \neq 0 \to$ inconsistent β_{OLS}

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'y$$

$$= (X'X)^{-1} X'(X\beta + \varepsilon)$$

$$= \cancel{(X'X)^{-1} X'X}\beta + (X'X)^{-1} X'\varepsilon$$

$$= \beta + \underbrace{\frac{COV(X, \varepsilon)}{VAR(X)}}_{\text{bias}}$$

bias > 0 if $cov(X, \varepsilon) > 0$

< 0 if $cov(X, \varepsilon) < 0$

$$Y = X\beta + \varepsilon \qquad / * Z'$$

$$Z'Y = Z'X\beta + Z'\varepsilon \to 0$$

$$\boxed{\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y}$$

both

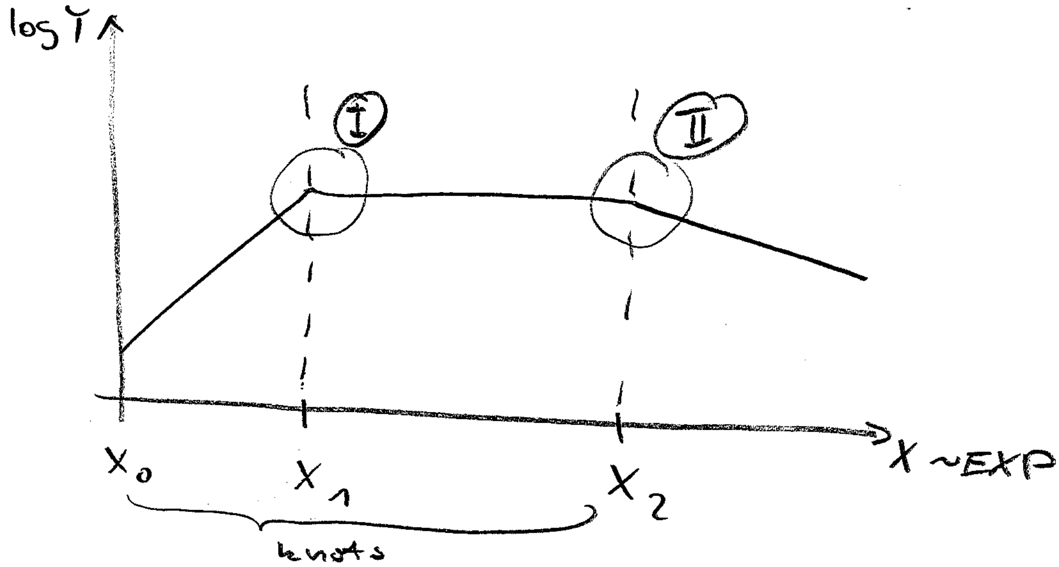
$cov(Z, \varepsilon) \overset{!}{=} 0$

$cov(Z, X) \neq 0$

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

$X_0, X_1, X_2 \sim$ knots $=$ const

$D_1 = 1$ if $X_0 \le X < X_1$

$D_2 = 1$ if $X_1 \le X < X_2$

$D_3 = 1$ if $X_2 < X$

$$Y = [a_1 + b_1(X - X_0)]D_1 + [a_2 + b_2(X - X_1)]D_2 + [a_3 + b_3(X - X_2)]D_3 + u$$

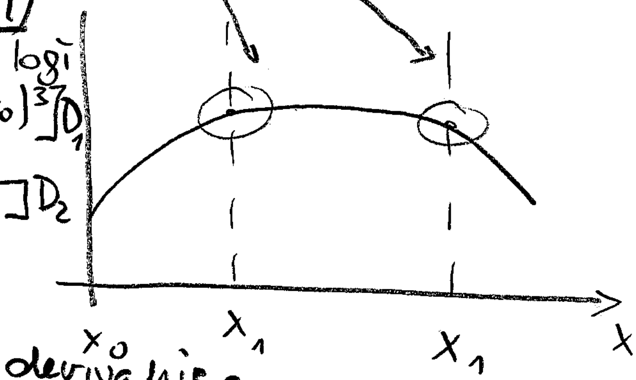
Impose continuity at I & II

$$a_2 = a_1 + b_1(X_1 - X_0)$$

$$a_3 = a_2 + b_2(X_2 - X_1)$$

- 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)

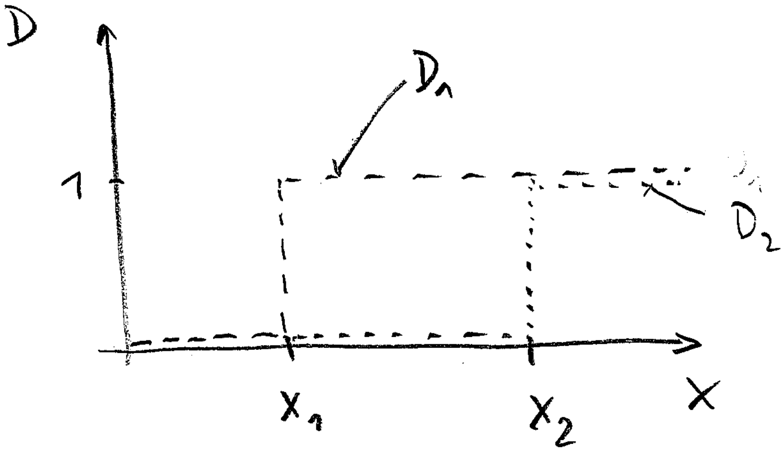
$$\boxed{Y = a_1 + b_1(X - X_0) + c_1(X - X_0)^2 + d_1(X - X_0)^3 + \sum_{i=1}^{k} (d_{i+1} - d_i)(X - X_i)^3 D_i^*}$$

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

marginal definit.

$k+1 = 3 \sim$ # intervals

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



- 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: