

AULA 28

REGRESSÃO

DESCONTÍNUA

Ernesto F. L. Amaral

20 de junho de 2013

Técnicas Avançadas de Avaliação de Políticas Públicas (DCP 098)

Fonte:

Curso “Técnicas Econométricas para Avaliação de Impacto” do “International Policy Centre for Inclusive Growth” (IPC-IG) da “United Nations Development Programme” (UNDP) (<http://www.ipc-undp.org/evaluation>).

INTRODUÇÃO

- Estamos interessados no efeito causal da variável de tratamento ($V.T. = D$) sobre uma variável dependente ($V.D. = Y$).
- D não é aleatoriamente distribuída.
- Porém, é conhecida uma variável X que, ao menos parcialmente, determina D .
- Regressão descontínua explora essa última relação.
- Por exemplo:
 - Recebimento de bolsa de estudo em universidade pode depender de teste de conhecimento.
 - Programa de transferência condicionada de renda (*conditional cash transfer program*, CCT) possui regras de elegibilidade (renda) e exige frequência à escola.

TIPOS DE REGRESSÃO DESCONTÍNUA

– Relação entre variável de tratamento (D) e variáveis observadas (X) pode ocorrer de duas formas.

1) **Determinística**: $D = 1\{X \geq c\}$, sendo que o ponto de corte (c) é conhecido.

– *Sharp Regression Discontinuity (SRD) Design.*

2) **Probabilístico**: D é uma variável aleatória dado determinadas características (X).

– Probabilidade condicional é descontínua em c.

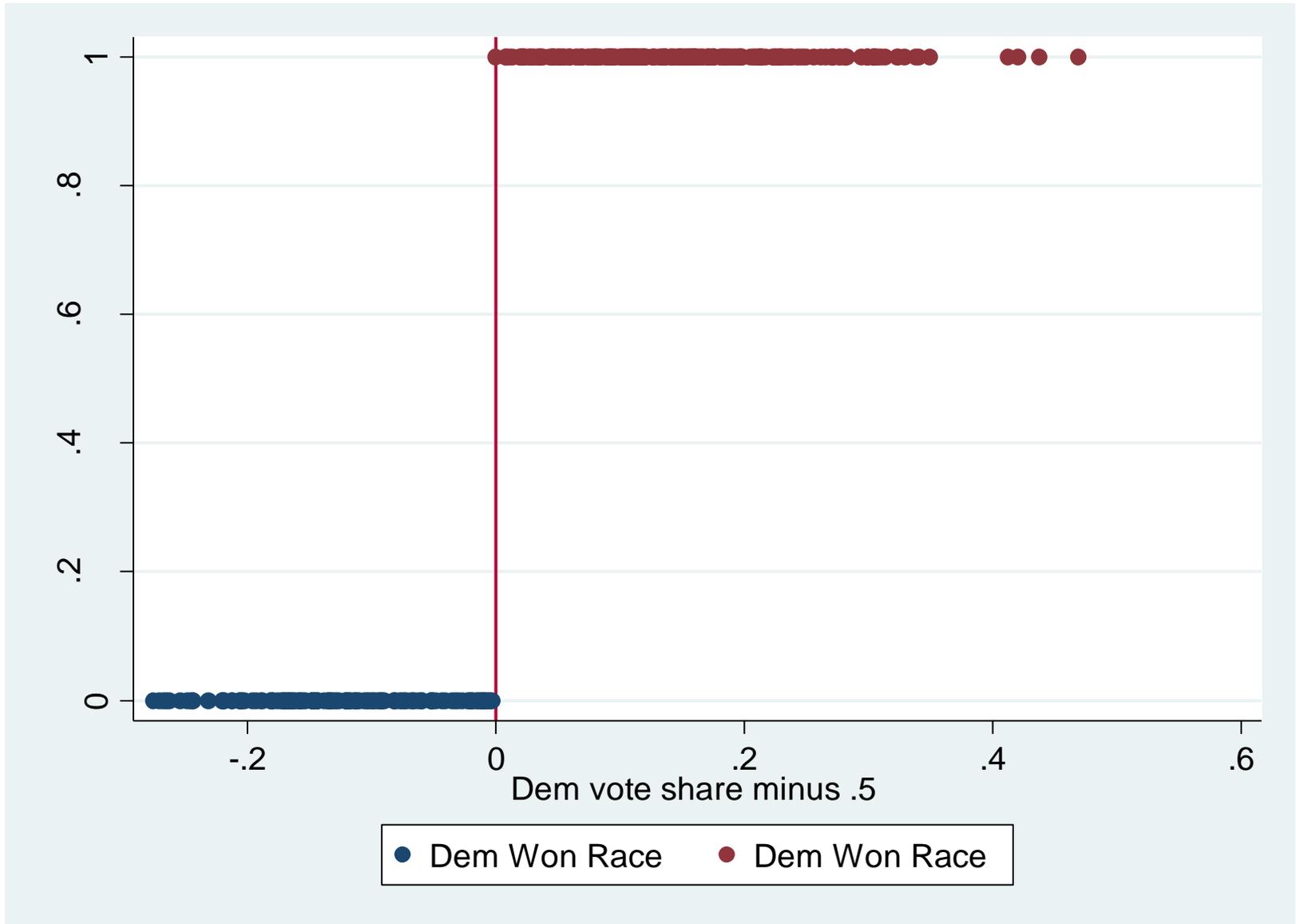
– $f(X) = E[D|X=x] = P(D=1|X=x)$.

– *Fuzzy Regression Discontinuity (FRD) Design.*

– Outras variáveis determinam o tratamento.

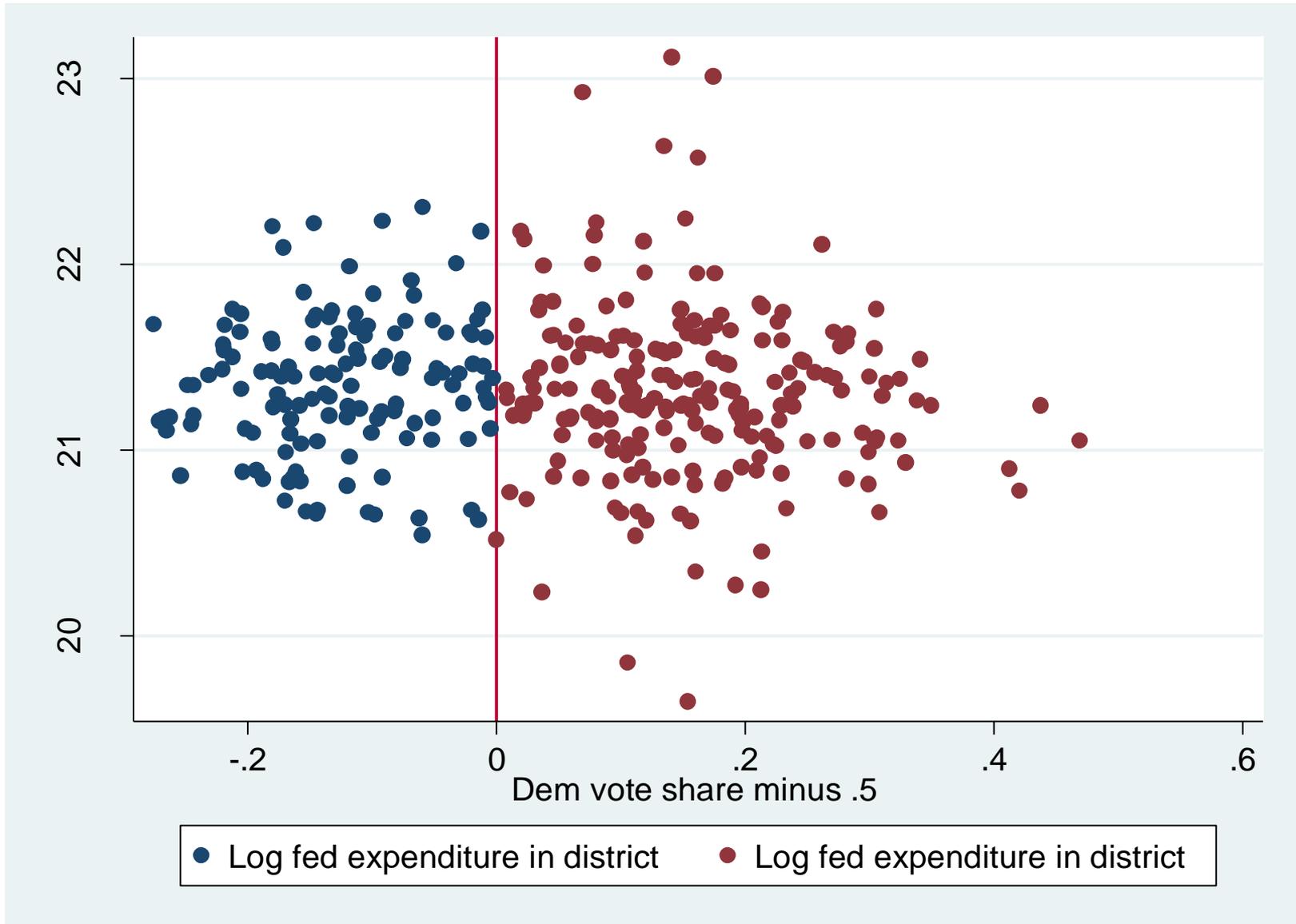
DETERMINÍSTICA (*SHARP*)

Variável de tratamento: democrata venceu (eixo vertical) por
Votação nos democratas menos 0,5 (eixo horizontal)



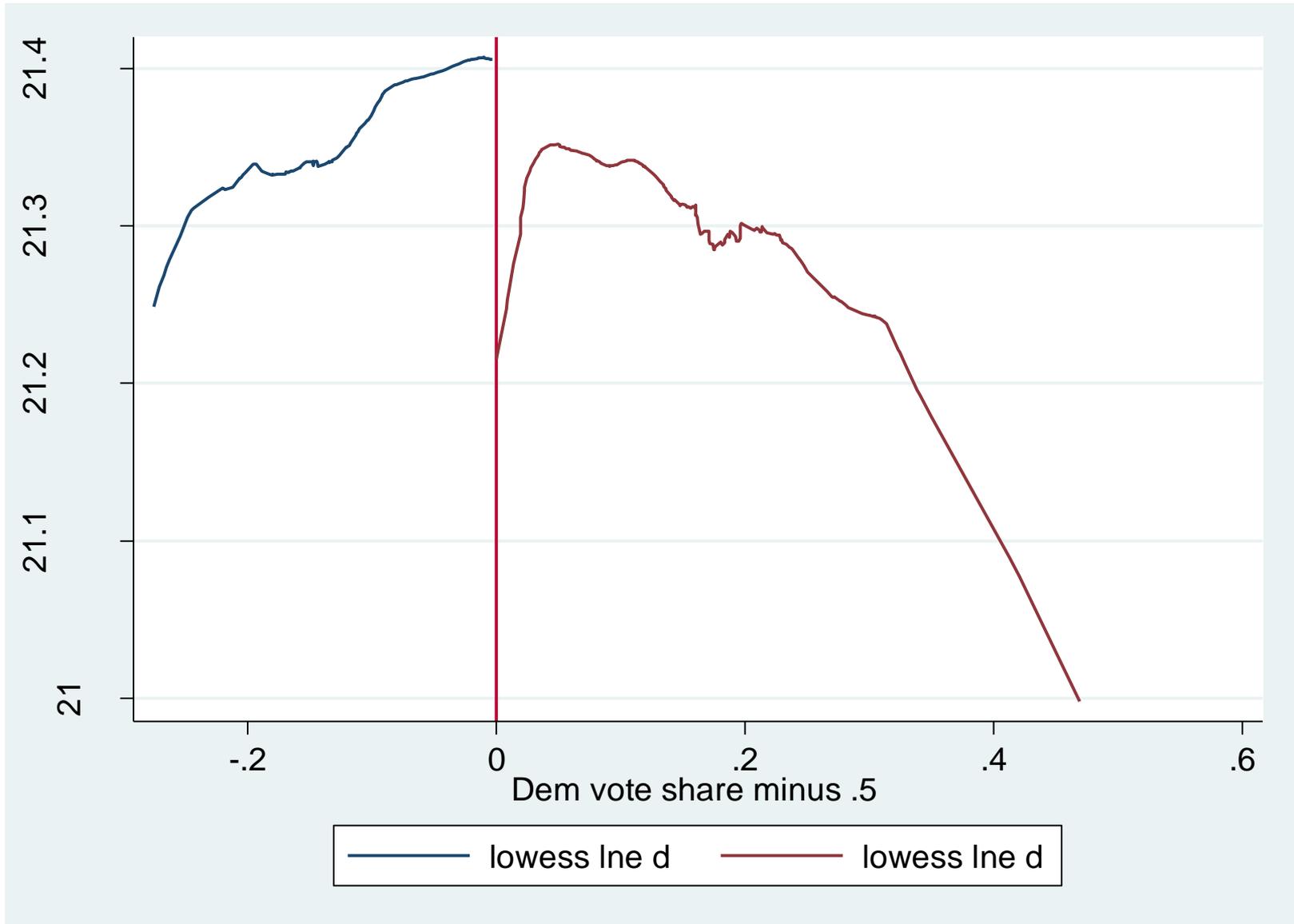
DETERMINÍSTICA (*SHARP*)

Variável dependente: gasto no distrito (eixo vertical) por
Votação nos democratas menos 0,5 (eixo horizontal)



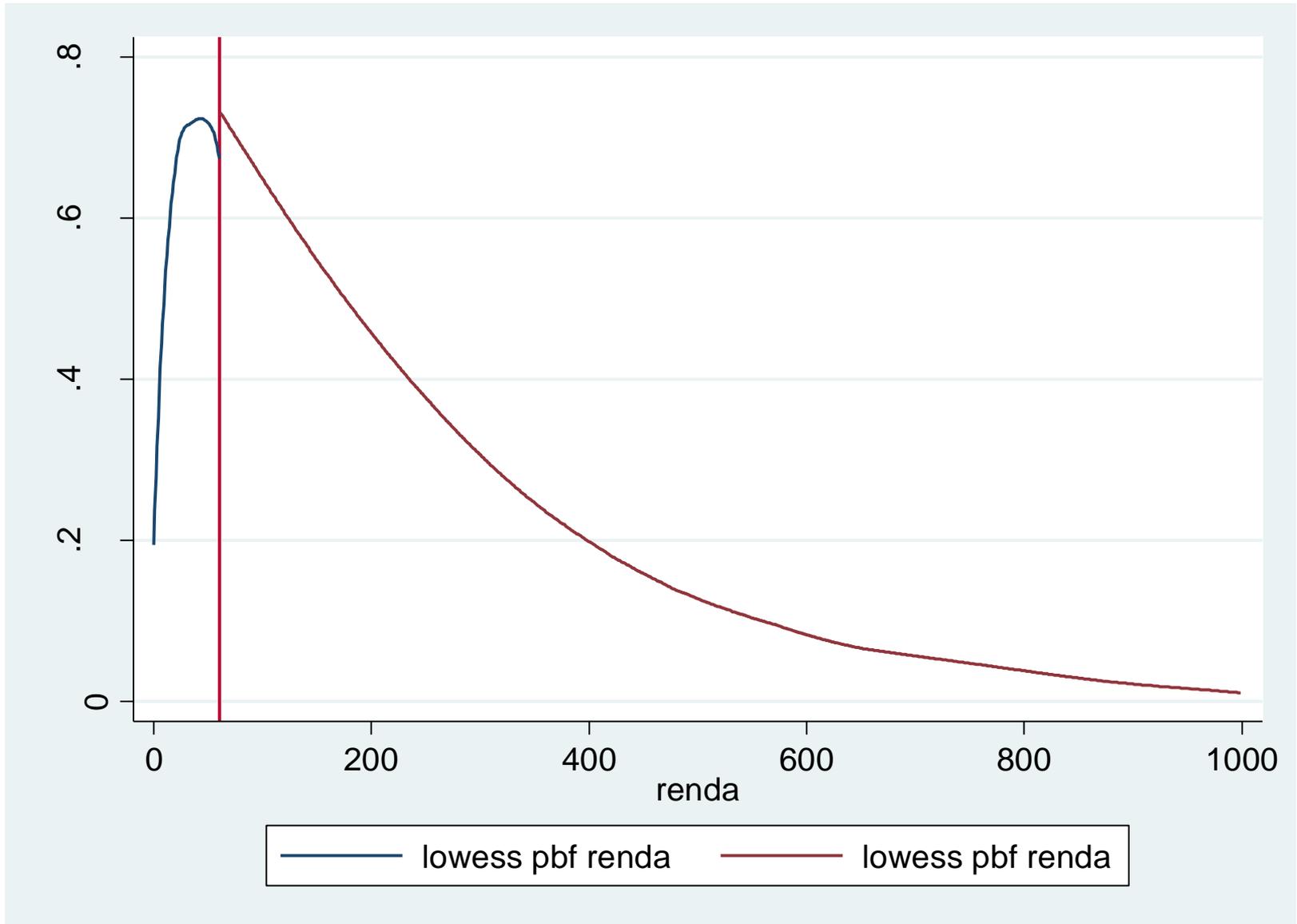
DETERMINÍSTICA (*SHARP*) SUAVIZADA (*LOWESS*)

Variável dependente: gasto no distrito (eixo vertical) por
Votação nos democratas menos 0,5 (eixo horizontal)



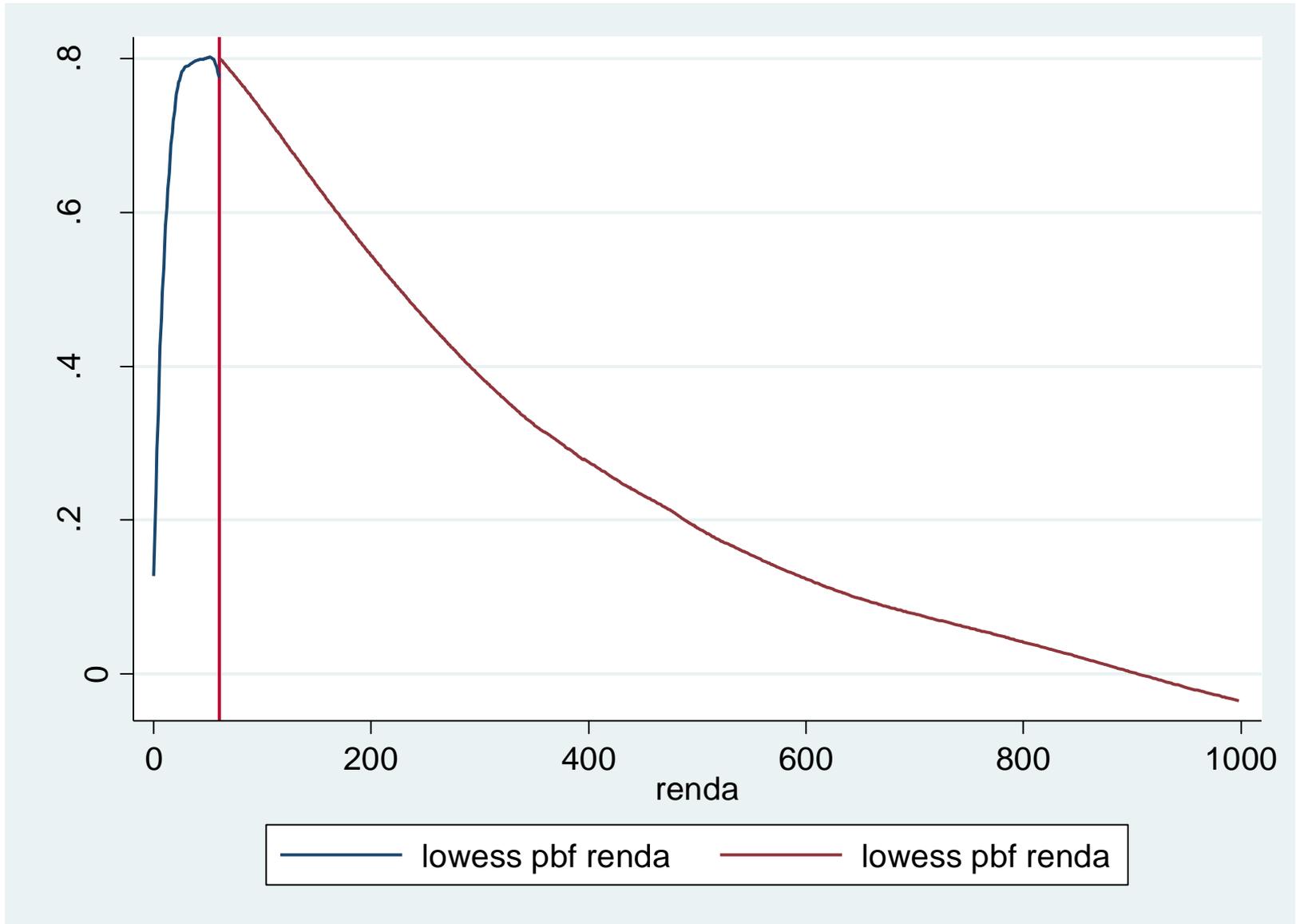
PROBABILÍSTICA (*FUZZY*)

V.T.: Percentual de domicílios com PBF (eixo vertical) por Corte de renda per capita de 60 reais (eixo horizontal)



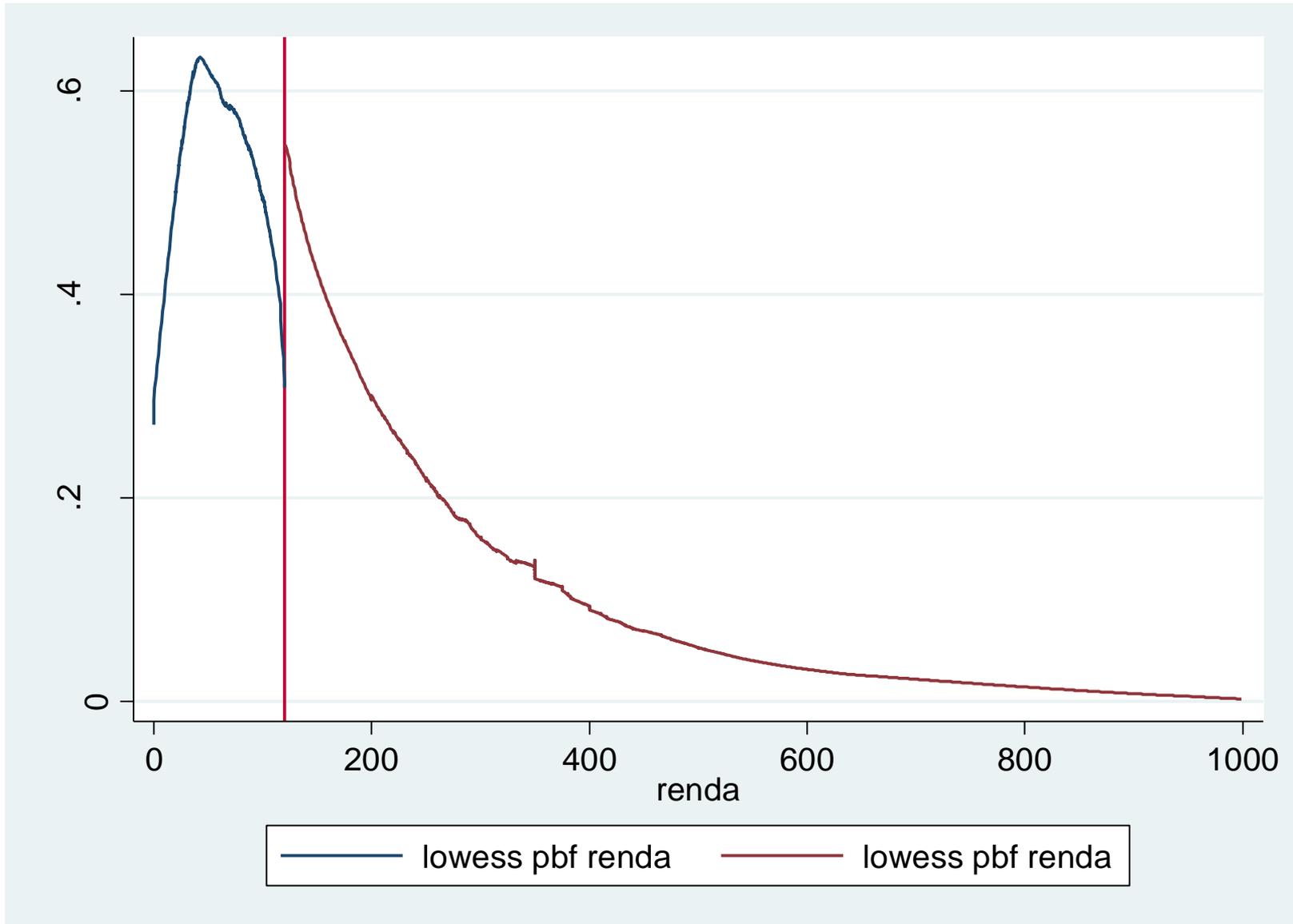
PROBABILÍSTICA (*FUZZY*)

V.T.: Percentual de indivíduos com PBF (eixo vertical) por Corte de renda per capita de 60 reais (eixo horizontal)



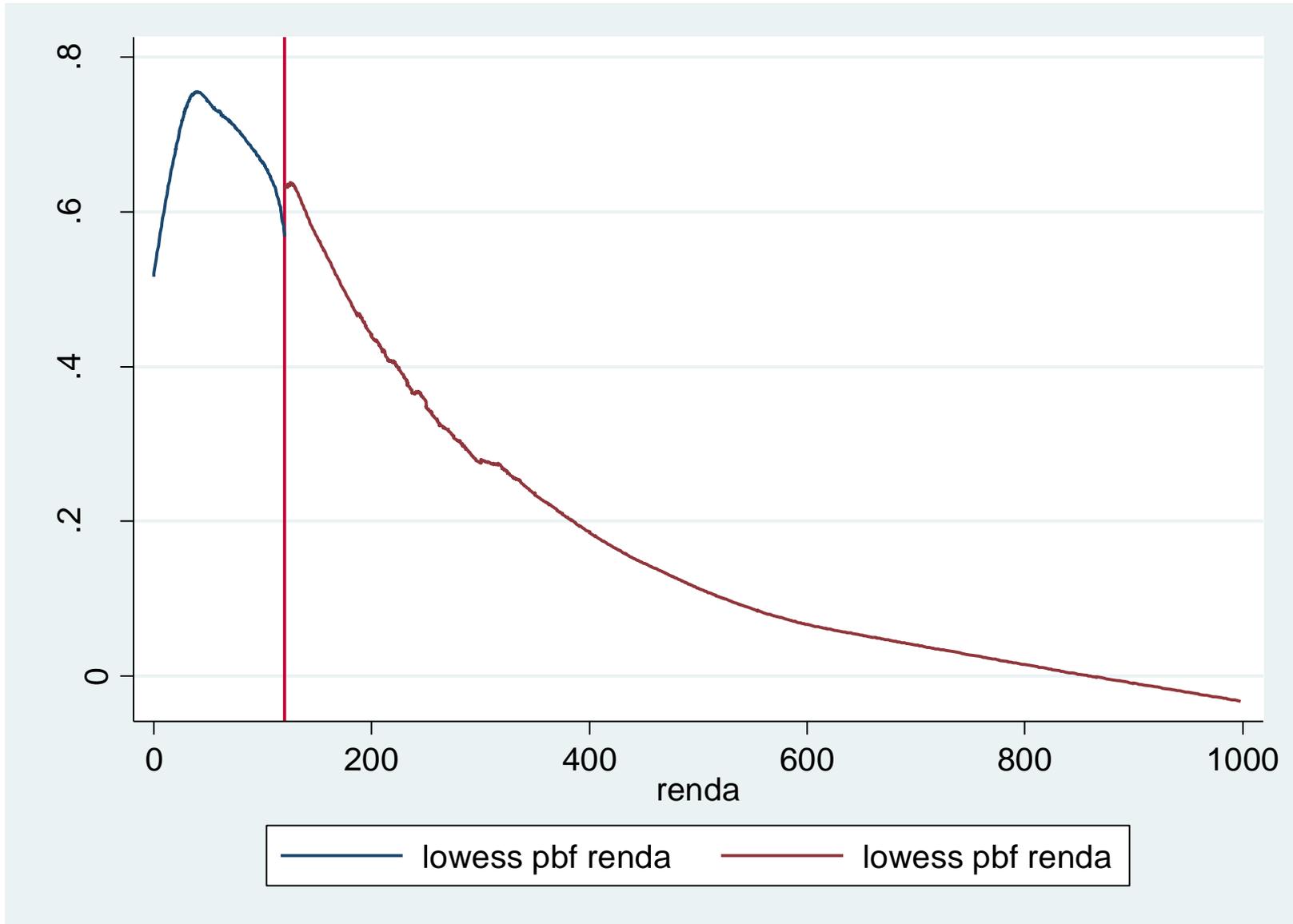
PROBABILÍSTICA (*FUZZY*)

V.T.: Percentual de domicílios com PBF (eixo vertical) por Corte de renda per capita de 120 reais (eixo horizontal)



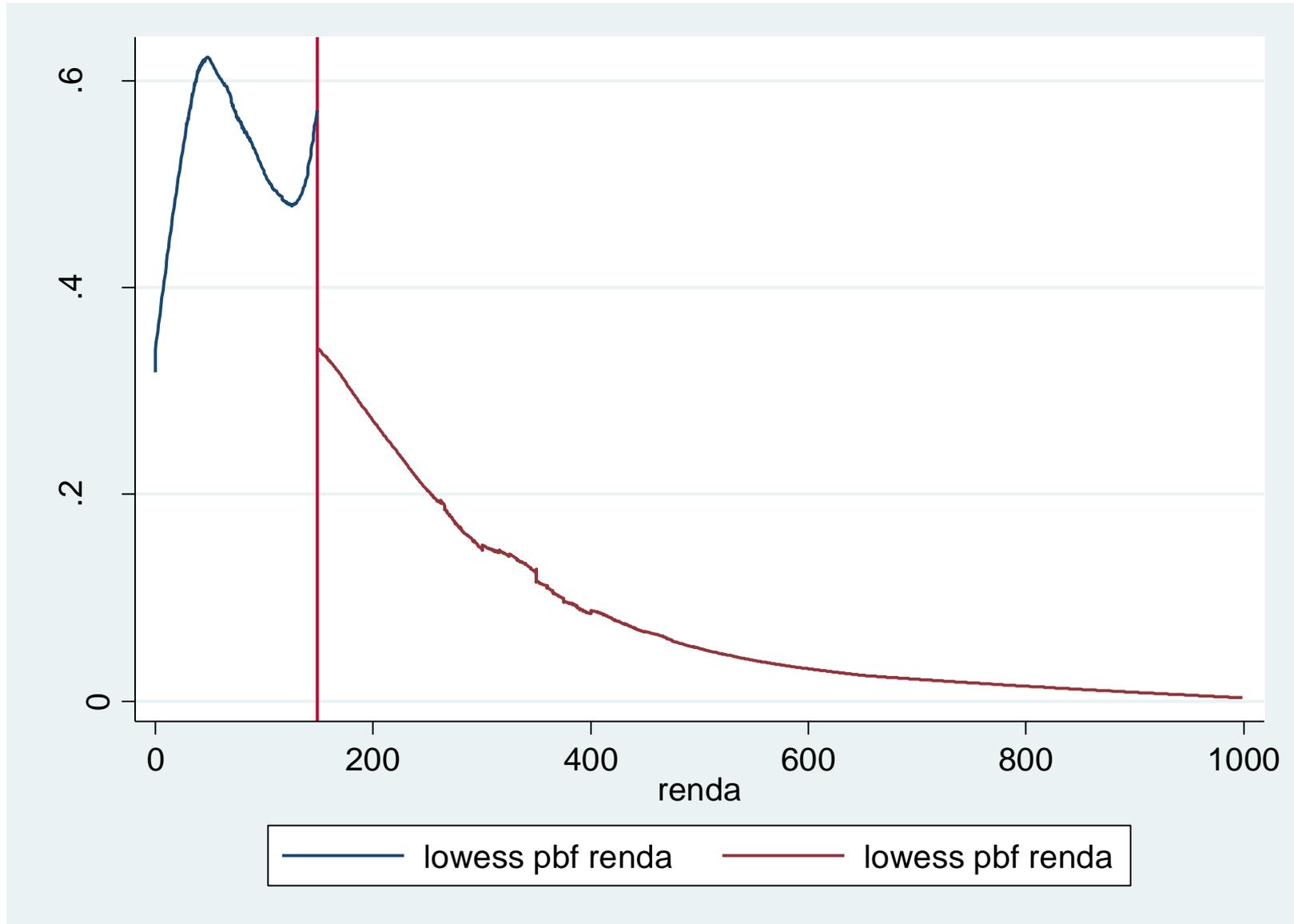
PROBABILÍSTICA (*FUZZY*)

V.T.: Percentual de indivíduos com PBF (eixo vertical) por Corte de renda per capita de 120 reais (eixo horizontal)



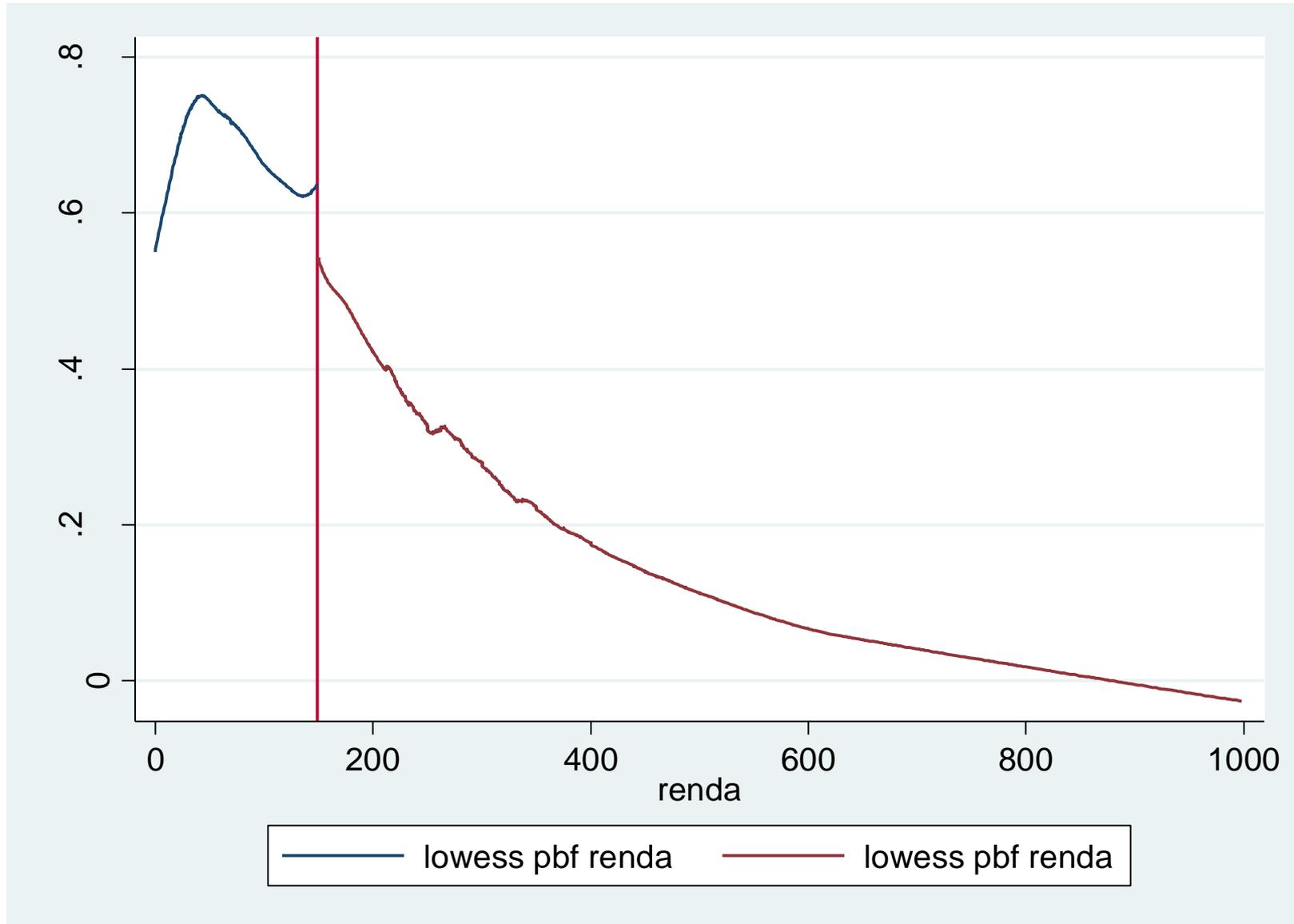
PROBABILÍSTICA (*FUZZY*)

V.T.: Percentual de domicílios com PBF (eixo vertical) por Corte de renda per capita de 149 reais (eixo horizontal)



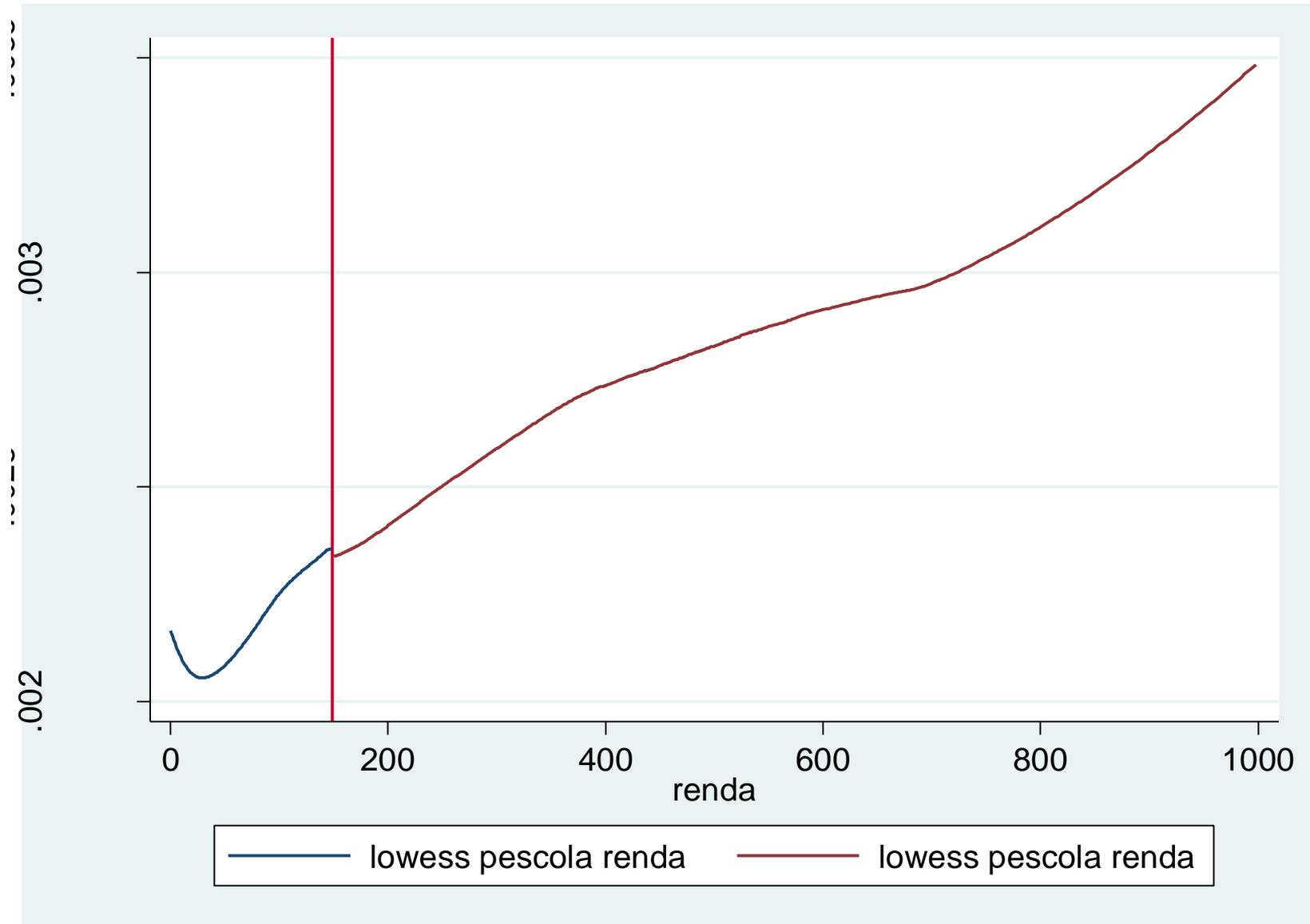
PROBABILÍSTICA (*FUZZY*)

V.T.: Percentual de indivíduos com PBF (eixo vertical) por Corte de renda per capita de 149 reais (eixo horizontal)



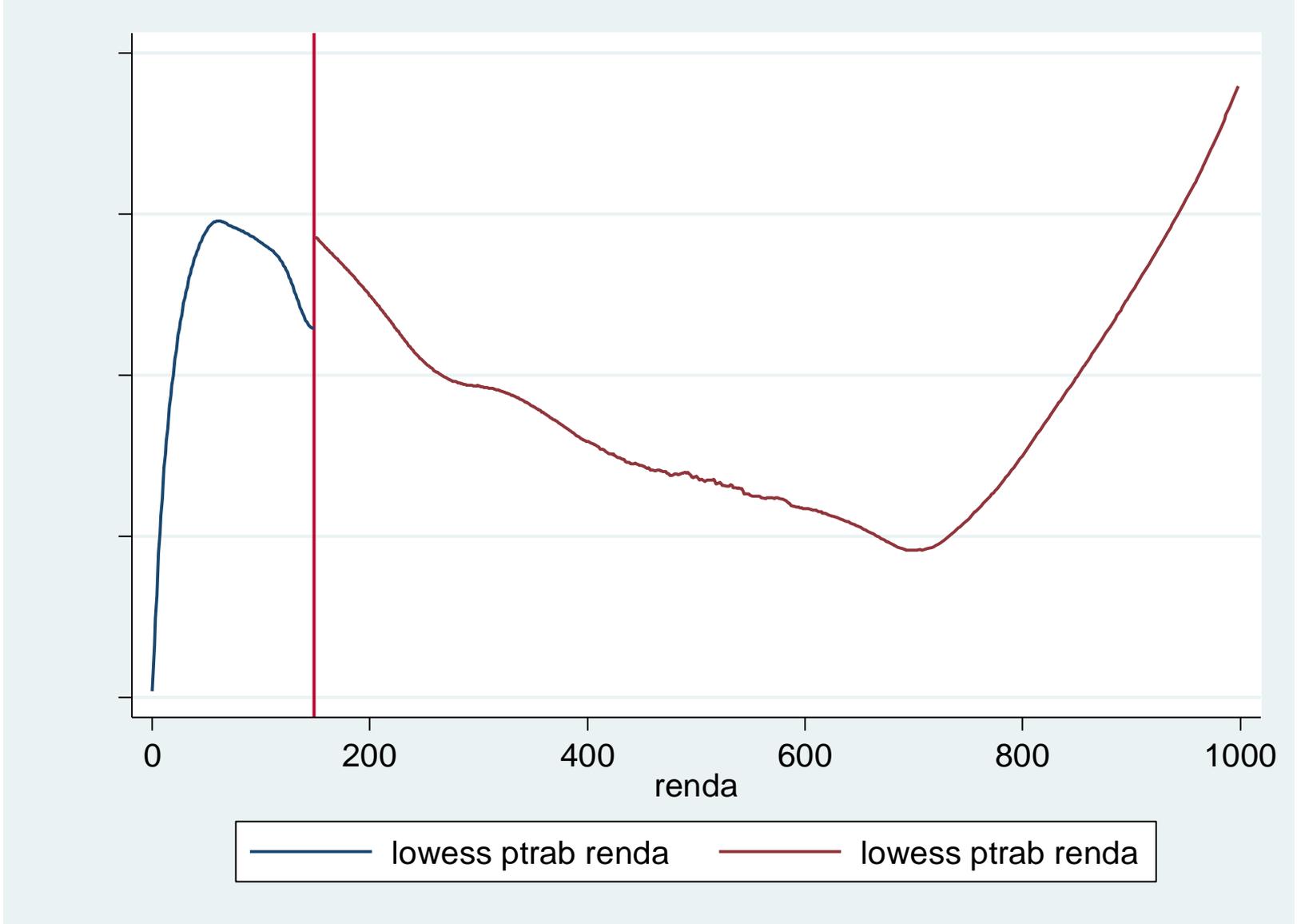
PROBABILÍSTICA (*FUZZY*)

V.D.: Percentual de indivíduos na escola (eixo vertical) por Corte de renda per capita de 149 reais (eixo horizontal)



PROBABILÍSTICA (*FUZZY*)

V.D.: Percentual de indivíduos no trabalho (eixo vertical) por Corte de renda per capita de 149 reais (eixo horizontal)



PROBLEMA FUNDAMENTAL DA AVALIAÇÃO

- Problema fundamental da avaliação é que não se observa $Y(1)$ para indivíduos com $X < c$, assim como não se observa $Y(0)$ para indivíduos com $X \geq c$.
- O que temos é:

$$E[Y|X] = \\ E[Y|D=0, X=x] * P(D=0|X=x) + \\ E[Y|D=1, X=x] * P(D=1|X=x)$$

***BANDWIDTH* (h)**

- O método da regressão descontínua depende da escolha do *bandwidth* (h).
- Quanto maior h, maior variação captada pelo estimador.
- No entanto, quanto maior h, menor a variância do estimador.
- Assim, deve haver um *bandwidth* ótimo.

ESCOLHENDO O *BANDWIDTH* (h)

- Não há consenso sobre como escolher h.
- Uma regra básica é utilizar $h = N^{1/5}$, em que N é o número de observações.
- Para SRD, Imbens e Lemiex (2007) propõem utilizar o valor de h que minimiza:

$$1/N \sum (Y - \mu(X))^2$$

- Para FRD, Imbens e Lemiex (2007) propõem o mesmo procedimento, calculando um h que minimiza o desvio em relação ao tratamento (D) e utilizando o valor mínimo dentre os dois mínimos.

EQUAÇÕES

- Conforme aulas passadas:

$$Y = Y_0(1 - D) + Y_1D$$

$$Y = Y_0 + (Y_1 - Y_0)D$$

$$Y = \alpha + \beta D$$

- Supondo um valor arbitrário de *bandwidth* pequeno (h):

$$E[Y | x = c + h] - E[Y | x = c - h] =$$

$$\{E[\alpha | x = c + h] - E[\alpha | x = c - h]\} +$$

$$\beta * \{E[D | x = c + h] - E[D | x = c - h]\}$$

- Dois pressupostos:

- β é constante.

- $E[\alpha | X = x]$ é contínuo em X no ponto c .

PRECAUÇÕES

- Precauções ao utilizar a regressão descontínua.
- Origem de descontinuidade:
 - Teste nas covariáveis.
 - Teste de continuidade.
- Escolha do *bandwidth*:
 - Teste de diferentes proporções do *bandwidth* escolhido.