

Variáveis Instrumentais (auto-seleção em não-observáveis)

- Como estimar o ATE quando

$$\Pr [T = 1 | Y(1), Y(0), X] \neq \Pr [T = 1 | X] = p(X)$$

- Na literatura econométrica, tem-se um problema equivalente:

$$Y_i = \alpha + \beta \cdot T_i + \epsilon_i$$

$$T_i = \mathbf{1} \{ \gamma + \delta \cdot Z_i + \eta_i \geq 0 \}$$

$$Cov(\eta_i, \epsilon_i) \neq 0, \quad Cov(Z_i, \epsilon_i) = 0$$

- A principal diferença está em que o modelo anterior impõe efeitos constantes $\beta_i = \beta \forall i$ e “single-index model” no primeiro estágio.
- Mostraremos aqui como interpretar e identificar β como sendo ATE para uma subpopulação particular. Precisamente, β será o ATE para aqueles indivíduos para quem um instrumento Z afeta apenas numa direção a probabilidade de tratamento. Esse parâmetro é chamado de $LATE$ (local average treatment effects).

- Considere, para efeitos de simplificação, que Z é binária (embora argumento funcione para instrumentos com múltiplos valores).
- Defina tratamento potencial $T_i(z)$, como o tratamento que indivíduo i teria se tivesse recebido $Z = z$. Para Z binário, tratamento observado simplifica para:

$$T_i = T_i(1) \cdot Z_i + T_i(0) \cdot (1 - Z_i)$$

- Resultado observado continua a ser:

$$Y_i = Y_i(1) \cdot T_i + Y_i(0) \cdot (1 - T_i)$$

- Hipóteses de identificação:

- Instrumento válido:

$$(Y(1), Y(0), T(1), T(0)) \perp\!\!\!\perp Z$$

- Monotonicidade:

$$T(1) \geq T(0)$$

- Essa hipótese impede que um dos quatro possíveis tipos de indivíduos exista (*defiers*):

$T(1)$	$T(0)$	T	tipo
1	1	1	<i>always-taker</i> (a)
1	0	se $Z = 1, T = 1$ se $Z = 0, T = 0$	<i>complier</i> (c)
0	1	se $Z = 1, T = 0$ se $Z = 0, T = 1$	<i>defier</i> (d)
0	0	0	<i>never-taker</i> (n)

- Pode-se mostrar que com essas hipóteses, identifica-se o efeito médio do tratamento para os “compliers”, $LATE$:

$$\begin{aligned}
 LATE &= E[Y(1) - Y(0) | T(1) > T(0)] \\
 &= \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[T|Z = 1] - E[T|Z = 0]}
 \end{aligned}$$

- Para ver porque isso ocorre, decomponha

$$\begin{aligned} E[Y|Z = 1] \\ &= E[Y|Z = 1, c] \cdot \Pr[c] + E[Y|Z = 1, a] \cdot \Pr[a] \\ &\quad + E[Y|Z = 1, n] \cdot \Pr[n] + E[Y|Z = 1, d] \cdot \Pr[d] \end{aligned}$$

mas

$$\begin{aligned} E[Y|Z = 1, a] &= E[Y(1)|Z = 1, a] = E[Y(1)|a] \\ E[Y|Z = 1, n] &= E[Y(0)|Z = 1, n] = E[Y(0)|n] \\ E[Y|Z = 1, c] &= E[Y(1)|Z = 1, c] = E[Y(1)|c] \\ E[Y|Z = 1, d] &= E[Y(0)|Z = 1, d] = E[Y(0)|d] \end{aligned}$$

- Analogamente:

$$\begin{aligned} & E [Y | Z = 0] \\ = & E [Y | Z = 0, c] \cdot \Pr [c] + E [Y | Z = 0, a] \cdot \Pr [a] \\ & + E [Y | Z = 0, n] \cdot \Pr [n] + E [Y | Z = 0, d] \cdot \Pr [d] \end{aligned}$$

mas

$$\begin{aligned} E [Y | Z = 0, a] &= E [Y (1) | Z = 0, a] = E [Y (1) | a] \\ E [Y | Z = 0, n] &= E [Y (0) | Z = 0, n] = E [Y (0) | n] \\ E [Y | Z = 0, c] &= E [Y (0) | Z = 0, c] = E [Y (0) | c] \\ E [Y | Z = 0, d] &= E [Y (1) | Z = 0, d] = E [Y (0) | d] \end{aligned}$$

- Portanto

$$\begin{aligned} E[Y|Z = 1] - E[Y|Z = 0] &= (E[Y(1)|c] - E[Y(0)|c]) \cdot \Pr[c] \\ &\quad - (E[Y(1)|d] - E[Y(0)|d]) \cdot \Pr[d] \\ &= E[Y(1) - Y(0)|c] \cdot \Pr[c] \end{aligned}$$

pois $\Pr[d] = 0$ (monotonicidade).

- Mas o que é $\Pr[c]$?

$$\begin{aligned} E[T|Z = 1] &= \Pr[T = 1|Z = 1] = \Pr[c] + \Pr[a] \\ E[T|Z = 0] &= \Pr[T = 1|Z = 0] = \Pr[a] + \Pr[d] \end{aligned}$$

Logo, como $\Pr[d] = 0$

$$\Pr[c] = E[T|Z = 1] - E[T|Z = 0]$$

- E finalmente

$$\begin{aligned}
 E [Y (1) - Y (0) |c] &= \frac{E [Y |Z = 1] - E [Y |Z = 0]}{\Pr [c]} \\
 &= \frac{E [Y |Z = 1] - E [Y |Z = 0]}{E [T |Z = 1] - E [T |Z = 0]} \\
 &= p \lim \left(\hat{\beta}_{2SLS} \right)
 \end{aligned}$$

onde $\hat{\beta}_{2SLS}$ é o estimador de dois estágios. No caso de instrumentos binários, $\hat{\beta}_{2SLS}$ é chamado de estimador de Wald:

$$\hat{\beta}_{2SLS} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{T}_1 - \bar{T}_0} = \left(\hat{\mathbf{T}}^\top \cdot \hat{\mathbf{T}} \right)^{-1} \cdot \hat{\mathbf{T}}^\top \cdot \mathbf{Y}$$

- Exemplos: (1) Guerra do Vietnam (Angrist) e (2) trimestre de nascimento (Angrist e Krueger)
 - Exemplo 1: como ir à guerra afeta renda e saúde (Y 's) após a guerra? Ir a guerra (T) é claramente endógeno. Angrist propôs usar o sorteio dos dias de nascimento como instrumento Z .
 - Exemplo 2: como educação (T) afeta renda (Y)? Educação endógena. Angrist e Krueger propõe usar trimestre de nascimento como instrumento para educação, sob argumento de que dada a regra de entrada na escola por trimestre de nascimento, escolaridade final poderia ser afetada.

Regressão com Descontinuidade (auto-seleção em não-observáveis)

- Novamente, o ideal seria estimar $ATE = E [Y (1) - Y (0)]$. Lembre-se que:

$$\begin{aligned} Y_i &= Y_i (0) + (Y_i (1) - Y_i (0)) \cdot T_i \\ &= \alpha + \beta \cdot T_i + Y_i (0) - \alpha + (Y_i (1) - Y_i (0) - \beta) \cdot T_i \\ &= \alpha + \beta \cdot T_i + U_i \end{aligned}$$

onde

$$\begin{aligned} \beta &= ATE \\ U_i &= Y_i (0) - \alpha + (Y_i (1) - Y_i (0) - \beta) \cdot T_i \end{aligned}$$

- Sabemos que o problema central na estimação de β está em que $Cov [U_i, T_i] \neq 0$.
- Mesmo adicionando covariadas X à regressão, $\hat{\beta}_{OLS}$ permanece inconsistente. Isto ocorre pois seleção ao tratamento depende de variáveis não observáveis.

- Suponha que exista uma variável contínua Z , tal que $E[T|Z]$ seja bem definida.
- Em um ponto de descontinuidade z_0 , a probabilidade de ser tratado dá um salto discreto:

$$T^+ \equiv \lim_{\epsilon \downarrow 0} \Pr [T = 1 | Z = z_0 + \epsilon]$$

$$T^- \equiv \lim_{\epsilon \downarrow 0} \Pr [T = 1 | Z = z_0 - \epsilon]$$

$$T^+ \neq T^-$$

- É esse salto discreto que será a fonte de variação exógena. Idéia é sermos capazes de identificar ATE não mais para toda a população, mas na vizinhança da descontinuidade.

- *Fuzzy design versus sharp design*

- sharp design: por exemplo,

$$Pr[T = 1|Z < z_0] = 0$$

$$Pr[T = 1|Z > z_0] = 1$$

- fuzzy design:

$$0 < Pr[T = 1|Z < z_0] < 1$$

$$0 < Pr[T = 1|Z > z_0] < 1$$

- Hipótese: continuidade da expectativa condicional dos resultados potenciais:

- $E[Y(1)|Z = z]$ e $E[Y(0)|Z = z]$ são contínuas em $z = z_0$.

- Defina

$$Y^+ \equiv \lim_{\epsilon \downarrow 0} E[Y|Z = z_0 + \epsilon]$$

$$Y^- \equiv \lim_{\epsilon \downarrow 0} E[Y|Z = z_0 - \epsilon]$$

- Exemplos: 2 exemplos importantes: Angrist e Lavy (tamanho de turma e proficiência) e van der Klaauw (bolsa e desempenho escolar na universidade)
- No primeiro exemplo, Angrist e Lavy exploram o fato de que em Israel a *Regra de Maimonides* determina que as classes em todas as escolas não possam passar de 40 alunos. Desta forma, podem-se comparar escolas, por exemplo, com 41 alunos em uma determinada série (e que devem ter formado duas turmas de 20 e 21) com escolas com um pouco menos de 40 alunos por série, as quais devem ter turmas únicas.

- No segundo exemplo, van der Klaauw explora a regra de oferta de bolsa baseada na nota de entrada do aluno. Se estiver acima de uma nota de corte, há oferta, caso contrário não há. Podem-se comparar alunos imediatamente acima com os imediatamente abaixo do *cutoff* em termos de desempenho ao longo da universidade para se ver qual o efeito de bolsa sobre desempenho.

- Definimos o ATE no ponto de descontinuidade z_0 como sendo

$$\begin{aligned} ATE(z_0) &= E [Y (1) - Y (0) | Z = z_0] \\ &= \frac{Y^+ - Y^-}{T^+ - T^-} \end{aligned}$$

isto ocorre pois

$$\begin{aligned} Y &= Y (0) + (Y (1) - Y (0)) \cdot T \\ \Rightarrow E [Y | Z = z_0 + \epsilon] &= E [Y (0) | Z = z_0 + \epsilon] \\ &\quad + E [(Y (1) - Y (0)) \cdot T | Z = z_0 + \epsilon] \end{aligned}$$

mas

$$\begin{aligned} Y^+ &= \lim_{\epsilon \downarrow 0} E [Y | Z = z_0 + \epsilon] \\ &= \lim_{\epsilon \downarrow 0} E [Y(0) | Z = z_0 + \epsilon] \\ &\quad + \lim_{\epsilon \downarrow 0} E [(Y(1) - Y(0)) \cdot T | Z = z_0 + \epsilon] \end{aligned}$$

e por analogia

$$\begin{aligned} Y^- &= \lim_{\epsilon \downarrow 0} E [Y | Z = z_0 - \epsilon] \\ &= \lim_{\epsilon \downarrow 0} E [Y(0) | Z = z_0 - \epsilon] \\ &\quad + \lim_{\epsilon \downarrow 0} E [(Y(1) - Y(0)) \cdot T | Z = z_0 - \epsilon] \end{aligned}$$

- Hipótese de identificação é que

$$E [Y (1) \cdot T | Z = z_0] = E [Y (1) | Z = z_0] \cdot E [T | Z = z_0]$$

$$E [Y (0) \cdot T | Z = z_0] = E [Y (0) | Z = z_0] \cdot E [T | Z = z_0]$$

ou mais fortemente

$$(Y (1), Y (0)) \perp\!\!\!\perp T | Z = z_0$$

- Portanto

$$\begin{aligned} Y^+ - Y^- &= \lim_{\epsilon \downarrow 0} E [Y (0) | Z = z_0 + \epsilon] - \lim_{\epsilon \downarrow 0} E [Y (0) | Z = z_0 - \epsilon] \\ &\quad + \lim_{\epsilon \downarrow 0} E [(Y (1) - Y (0)) \cdot T | Z = z_0 + \epsilon] \\ &\quad - \lim_{\epsilon \downarrow 0} E [(Y (1) - Y (0)) \cdot T | Z = z_0 - \epsilon] \\ &= \lim_{\epsilon \downarrow 0} E [Y (1) - Y (0) | Z = z_0 + \epsilon] \cdot T^+ \\ &\quad - \lim_{\epsilon \downarrow 0} E [Y (1) - Y (0) | Z = z_0 - \epsilon] \cdot T^- \\ &= E [Y (1) - Y (0) | Z = z_0] \cdot (T^+ - T^-) \end{aligned}$$

- É por isso que

$$\begin{aligned}ATE(z_0) &= E[Y(1) - Y(0) | Z = z_0] \\ &= \frac{Y^+ - Y^-}{T^+ - T^-}\end{aligned}$$

- Estimação segue por regressão polinomial (linear) local. Por exemplo:

$$\begin{aligned} & (a^+, b^+) \\ &= \arg \min_{a,b} \sum_{i=1}^N \mathbf{1}\{Z_i - z_0 > 0\} \cdot (Y_i - a - b \cdot (Z_i - z_0))^2 \cdot K_h(Z_i - z_0) \end{aligned}$$

onde

$$\widehat{Y}^+ = a^+$$

- Nesse caso, regressão não-paramétrica é importante a fim de se capturar salto à esquerda e à direita da descontinuidade
- Escolha de regressão polinomial tem a ver com fato de convergência ser mais rápida quanto mais termos polinomiais adicionarmos. Regressão por kernel, a qual corresponde ao polinômio de grau zero, não é adequada para estimação próximo aos limites.
- Cálculo do erro-padrão pode ser feito tal como proposto em Hahn, Todd e van der Klaauw (2001).

“Checklist” da implantação de RD (Lee e Lemieux)

1. Para checar se houve manipulação, mostre a distribuição da variável Z .
2. Apresente os principais gráficos de RD usando médias locais.
3. Caso use uma especificação polinomial, faça o gráfico.
4. Mostre sensibilidade dos resultados a diversas escolhas de h (bandwidth) on ordem do polinômio escolhido.
5. Conduza um RD em paralelo para as covariadas.
6. Mostre sensibilidade dos resultados à introdução de covariadas.

Figure 1: Simple Linear RD Setup

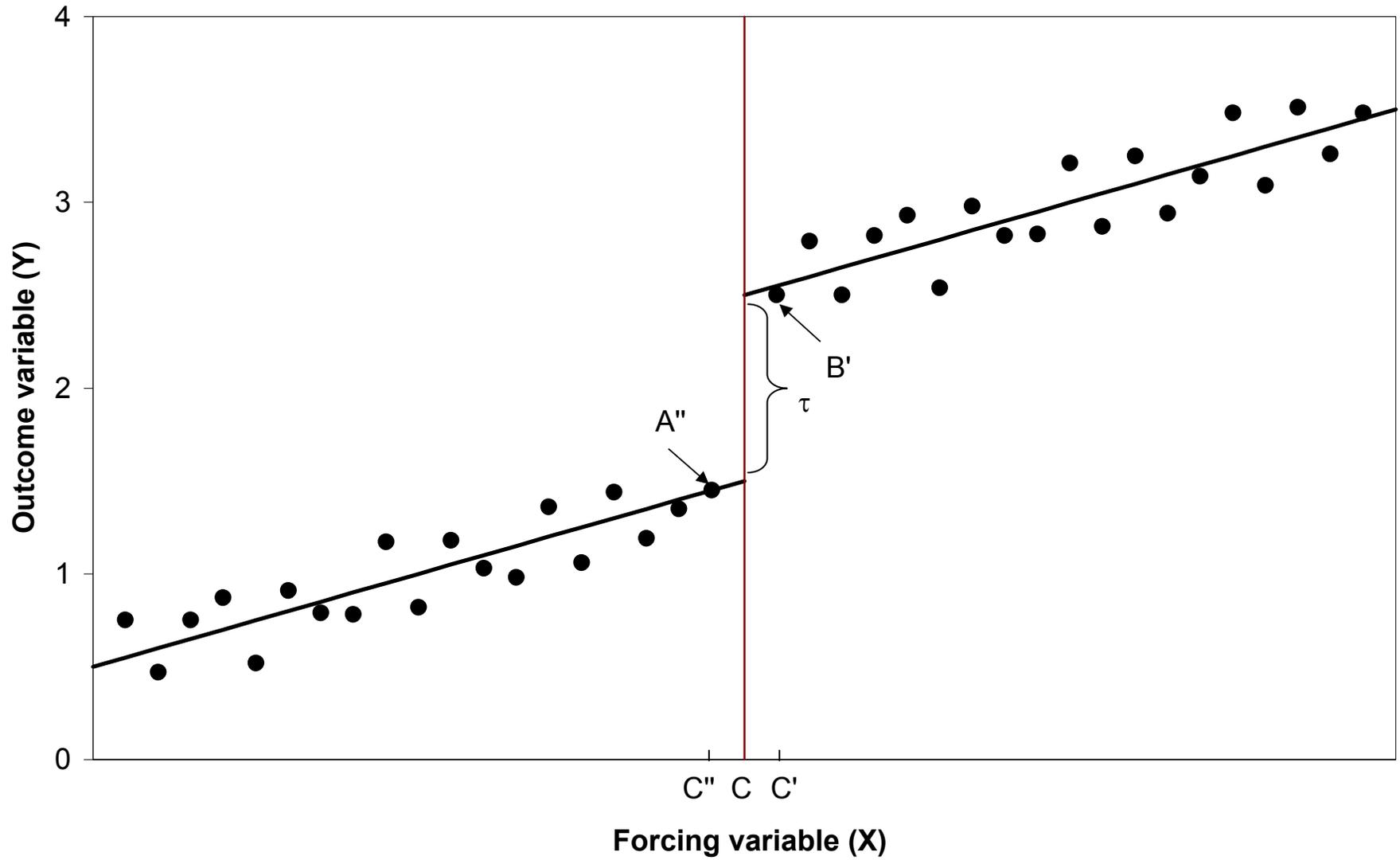


Figure 2: Nonlinear RD

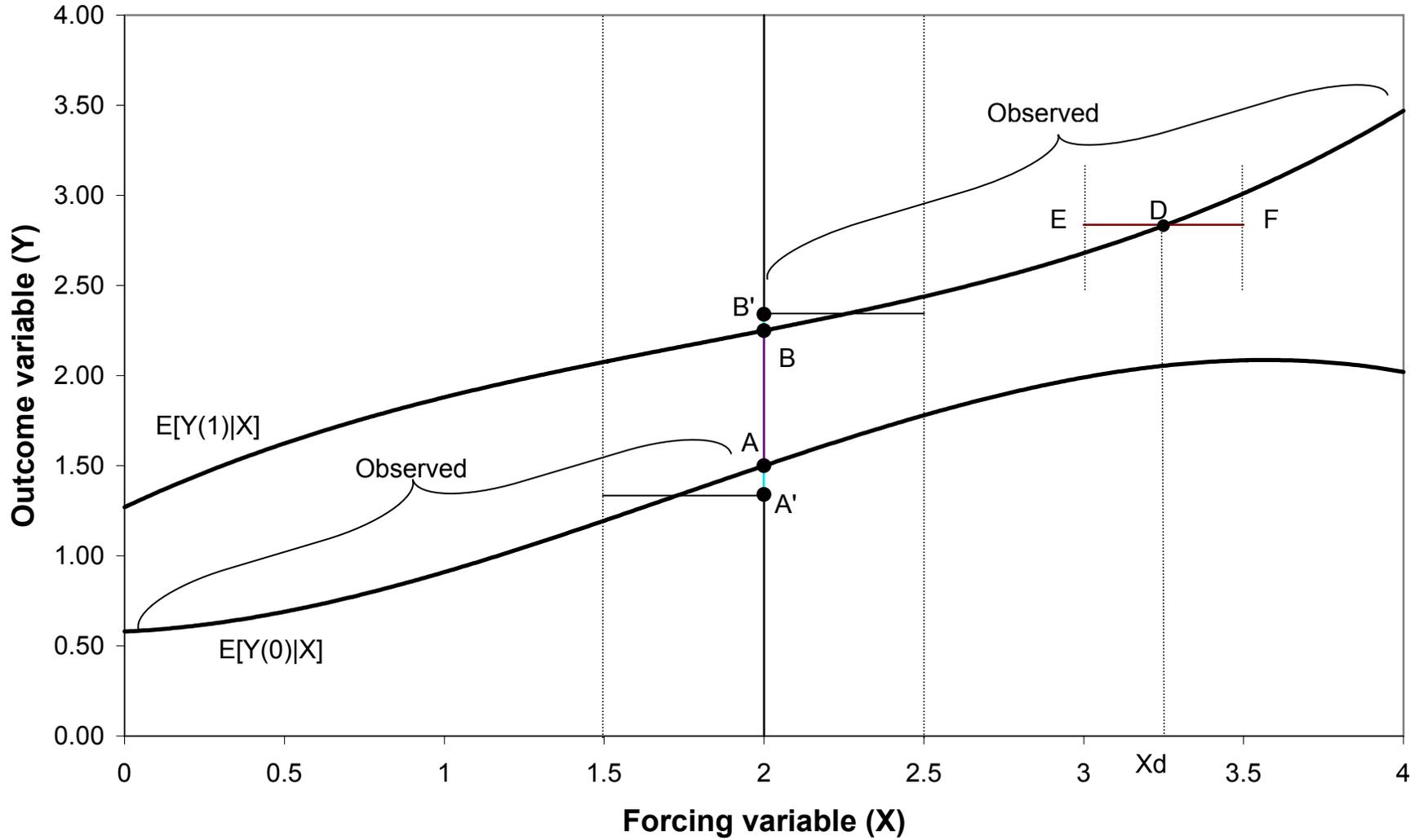


Figure 3: Randomized Experiment as a RD Design

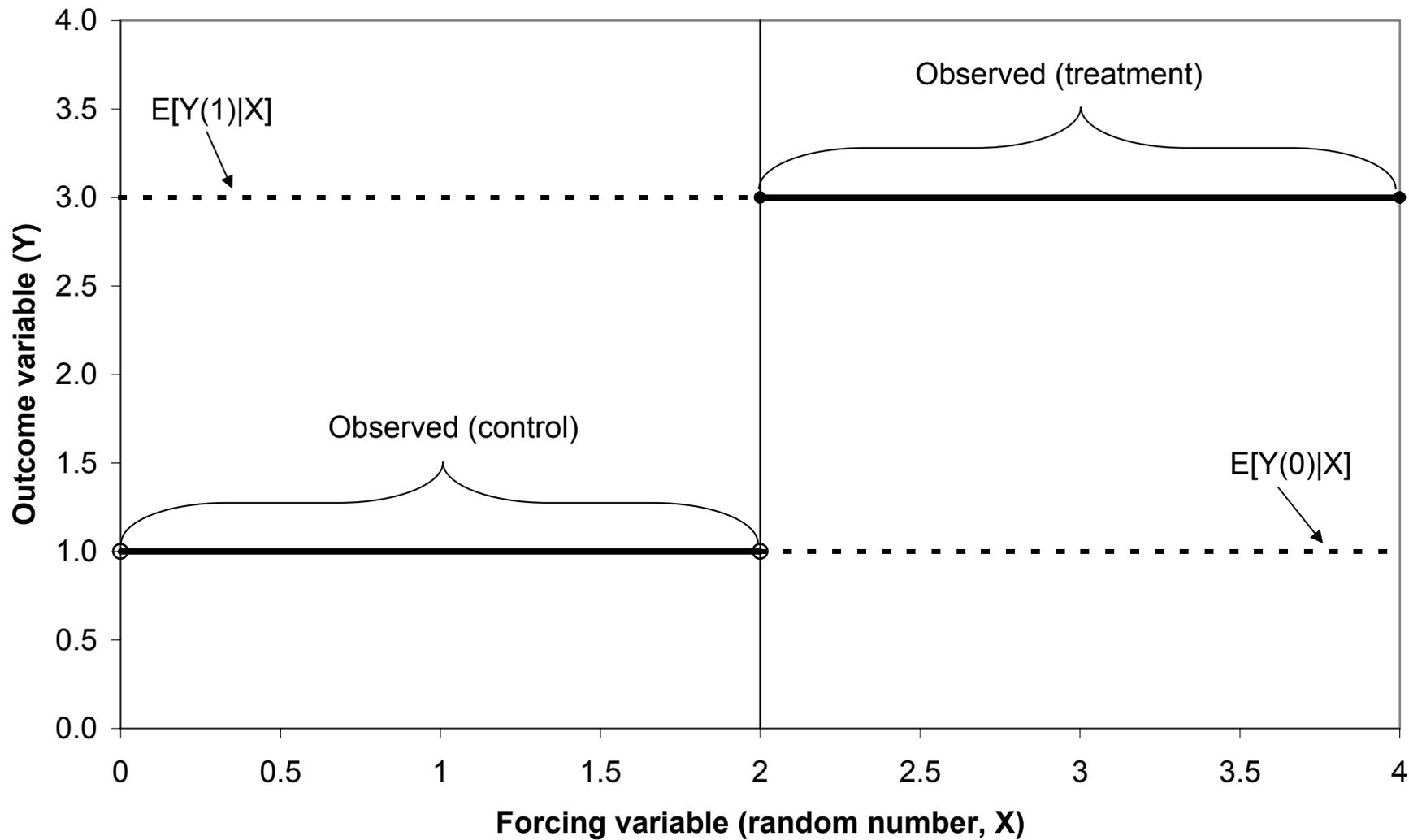


Figure 4: Density of Forcing Variable Conditional on $W=w, U=u$

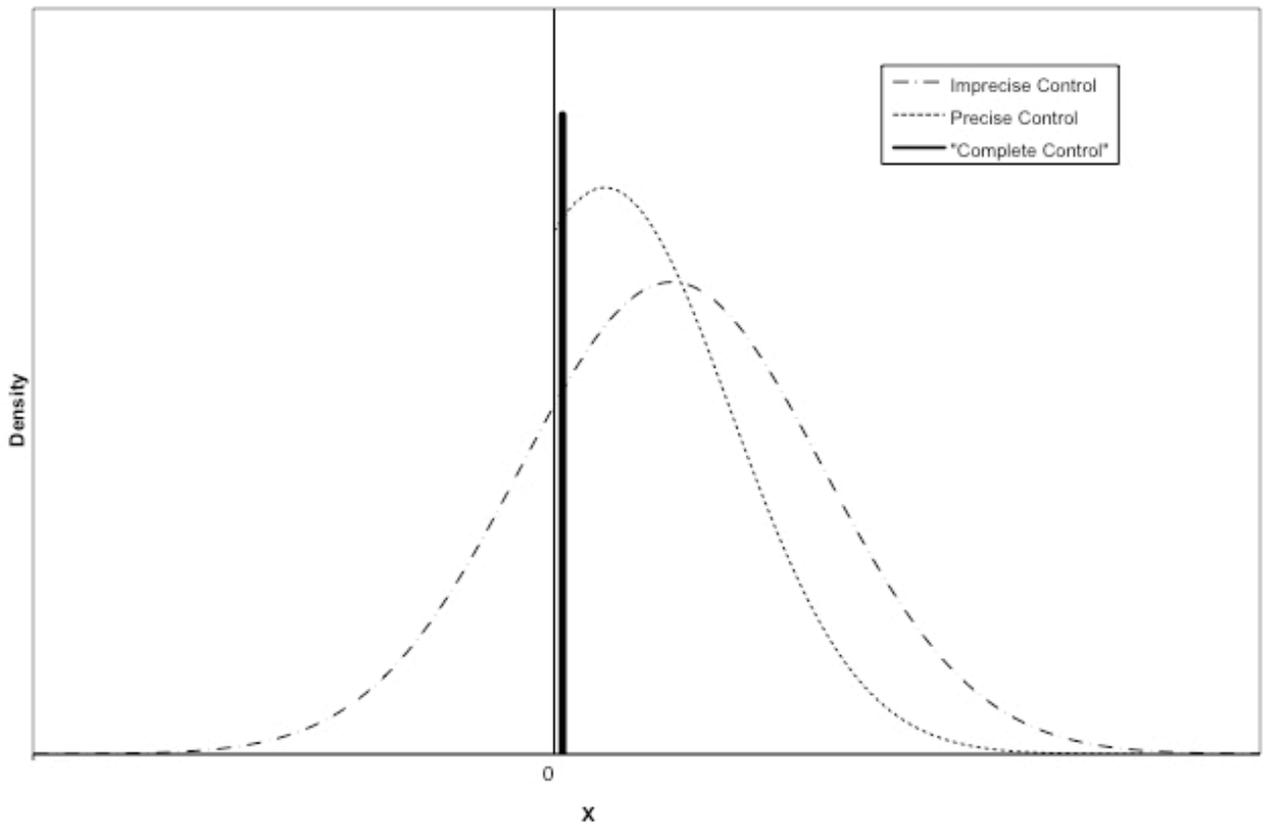
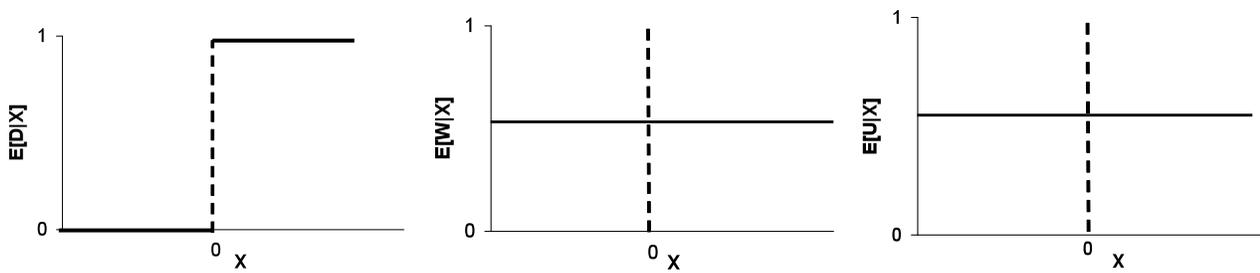
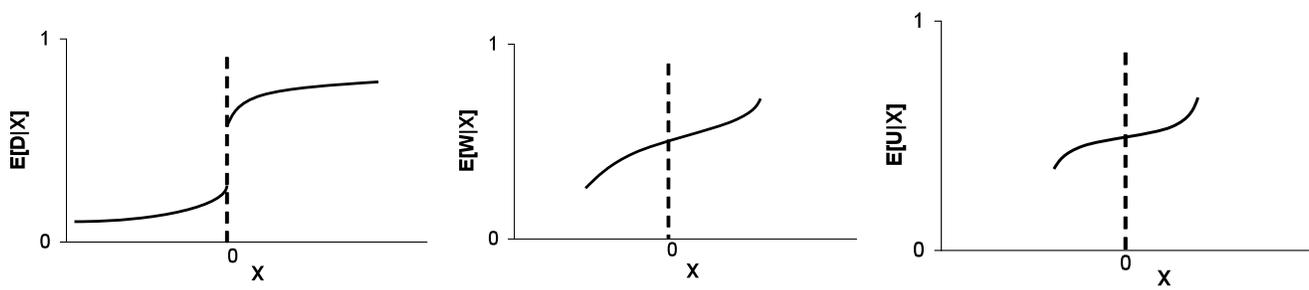


Figure 5. Treatment, Observables, and Unobservables in four research designs.

A. Randomized Experiment



B. Regression Discontinuity Design



C. Matching on Observables



D. Instrumental Variables

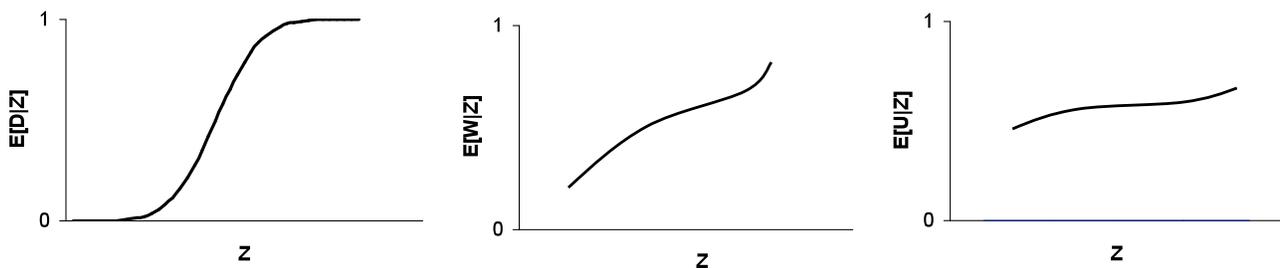
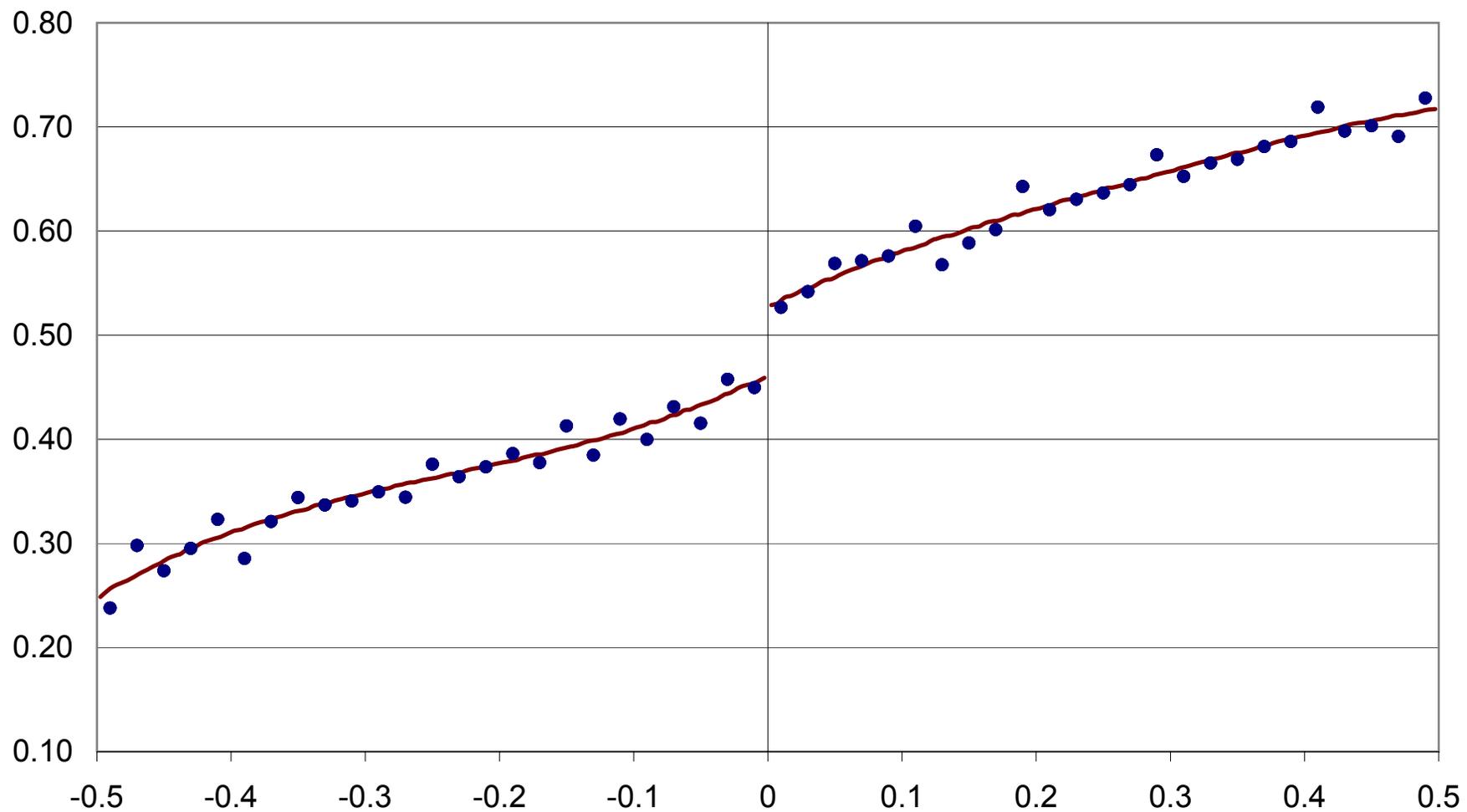
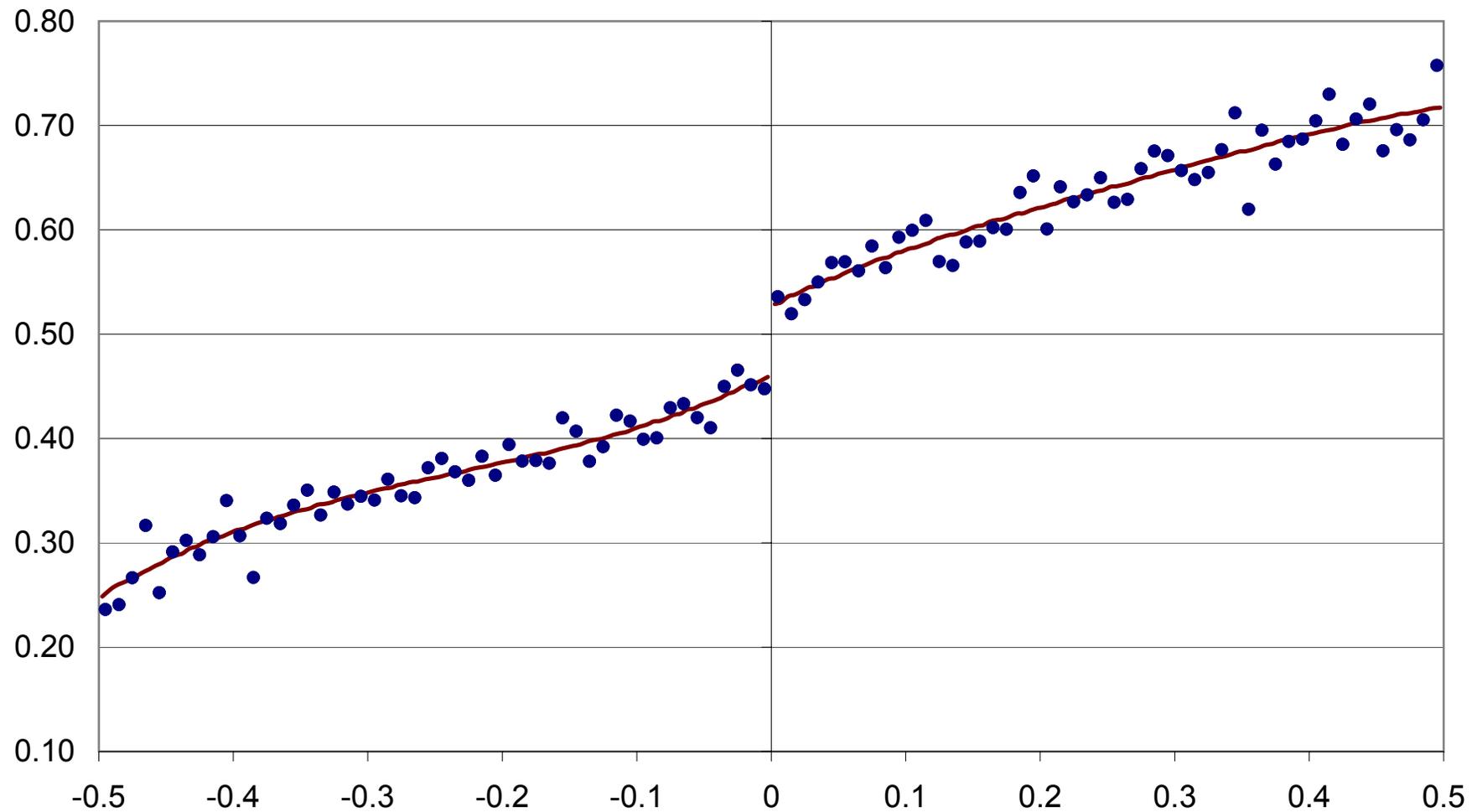


Figure 6a: Share of vote in next election, bandwidth of 0.02 (50 bins)



**Figure 6b: Share of vote in next election, bandwidth of 0.01
(100 bins)**



**Figure 6c: Share of vote in next election, bandwidth of 0.005
(200 bins)**

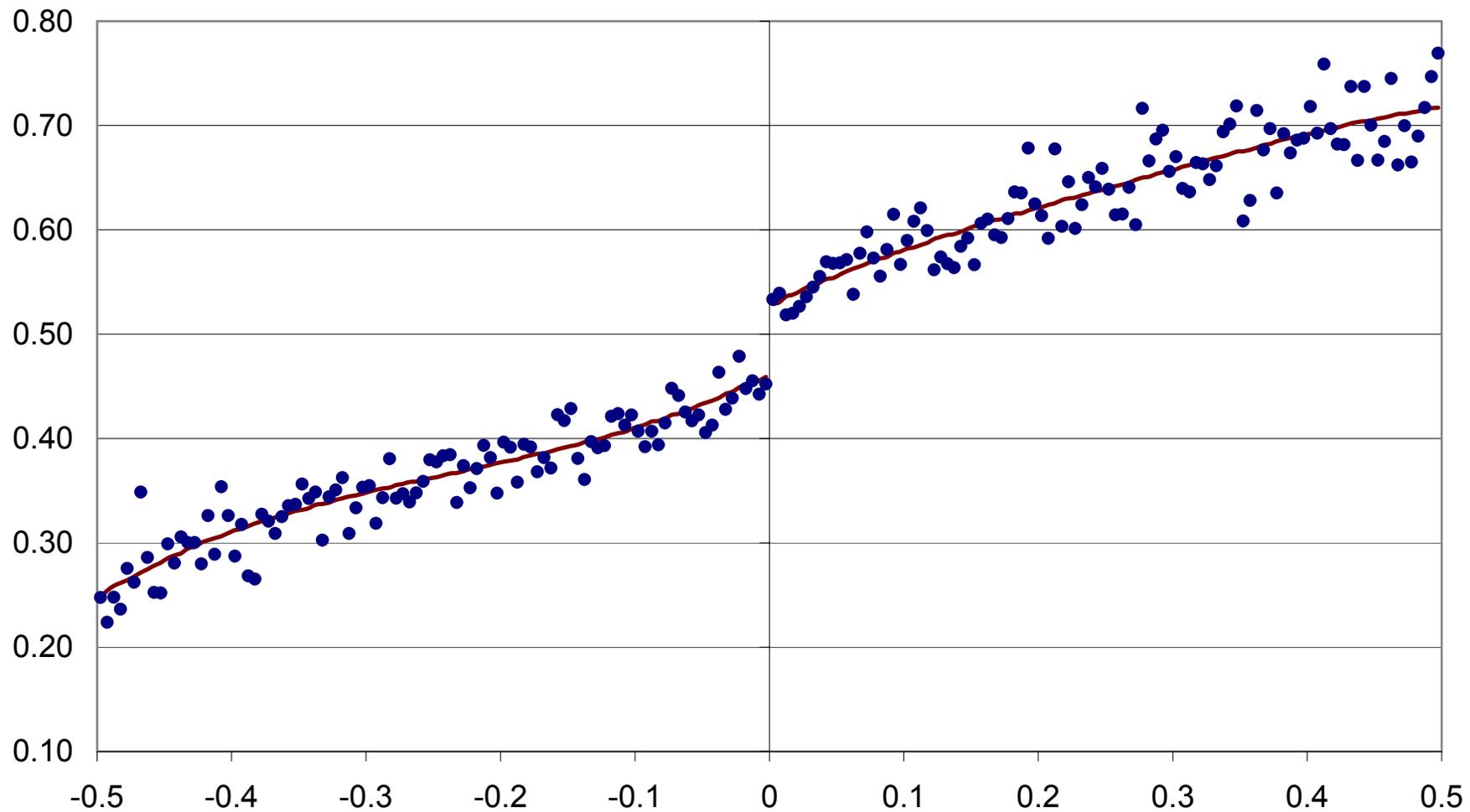


Figure 7a: Winning the next election, bandwidth of 0.02 (50 bins)

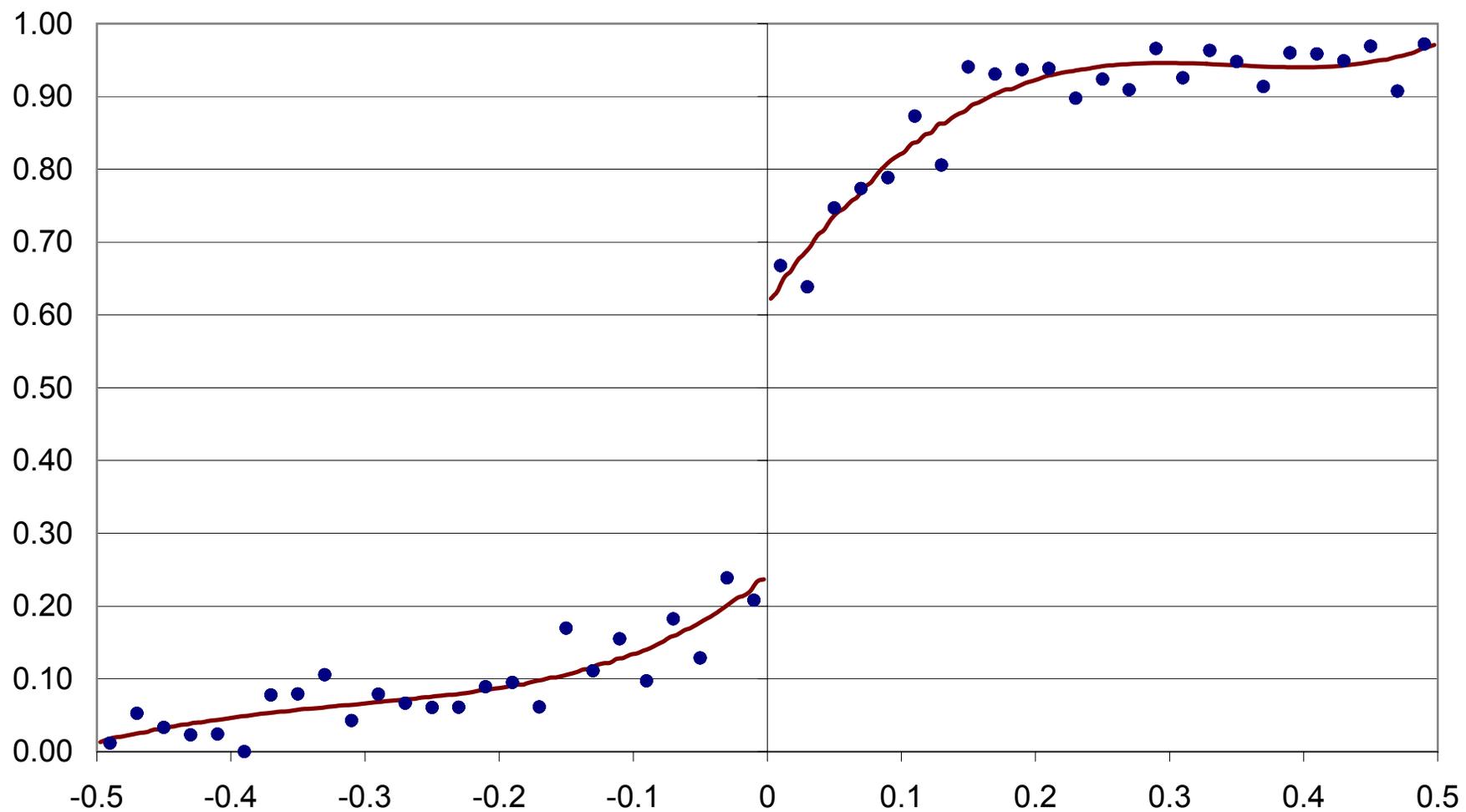


Figure 7b: Winning the next election, bandwidth of 0.01 (100 bins)

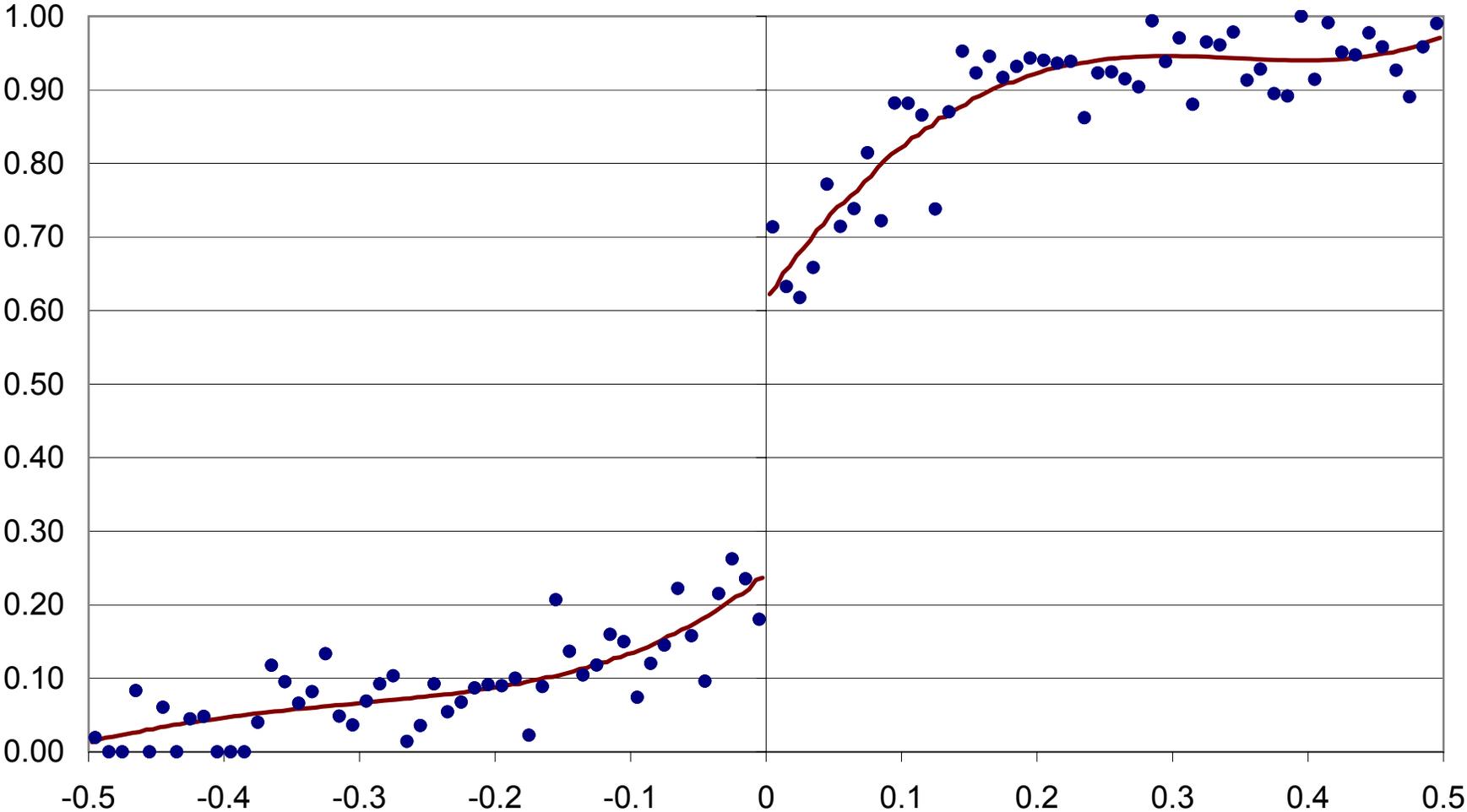


Figure 7c: Winning the next election, bandwidth of 0.005 (200 bins)

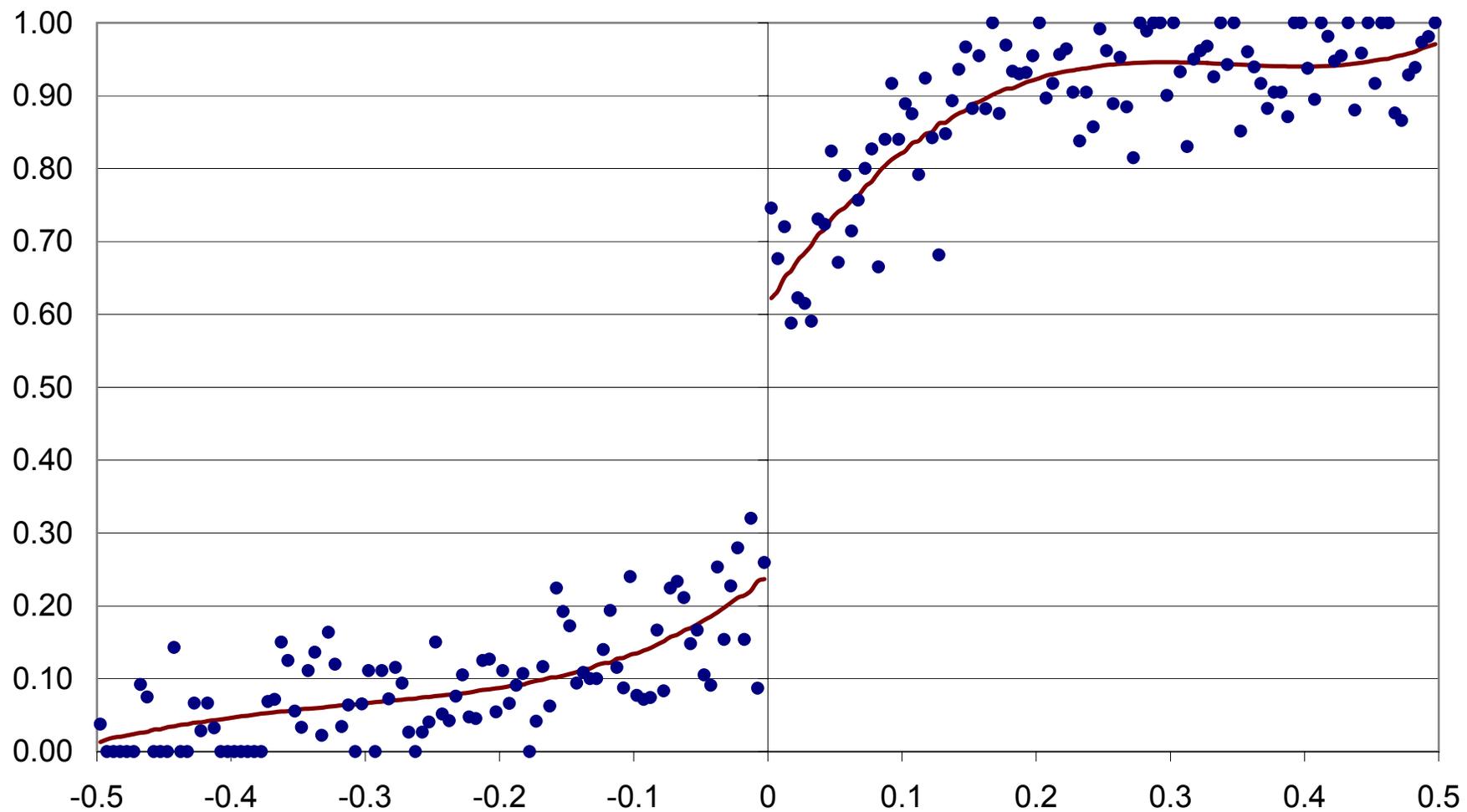


Figure 8: Density of the forcing variable (vote share in previous election)

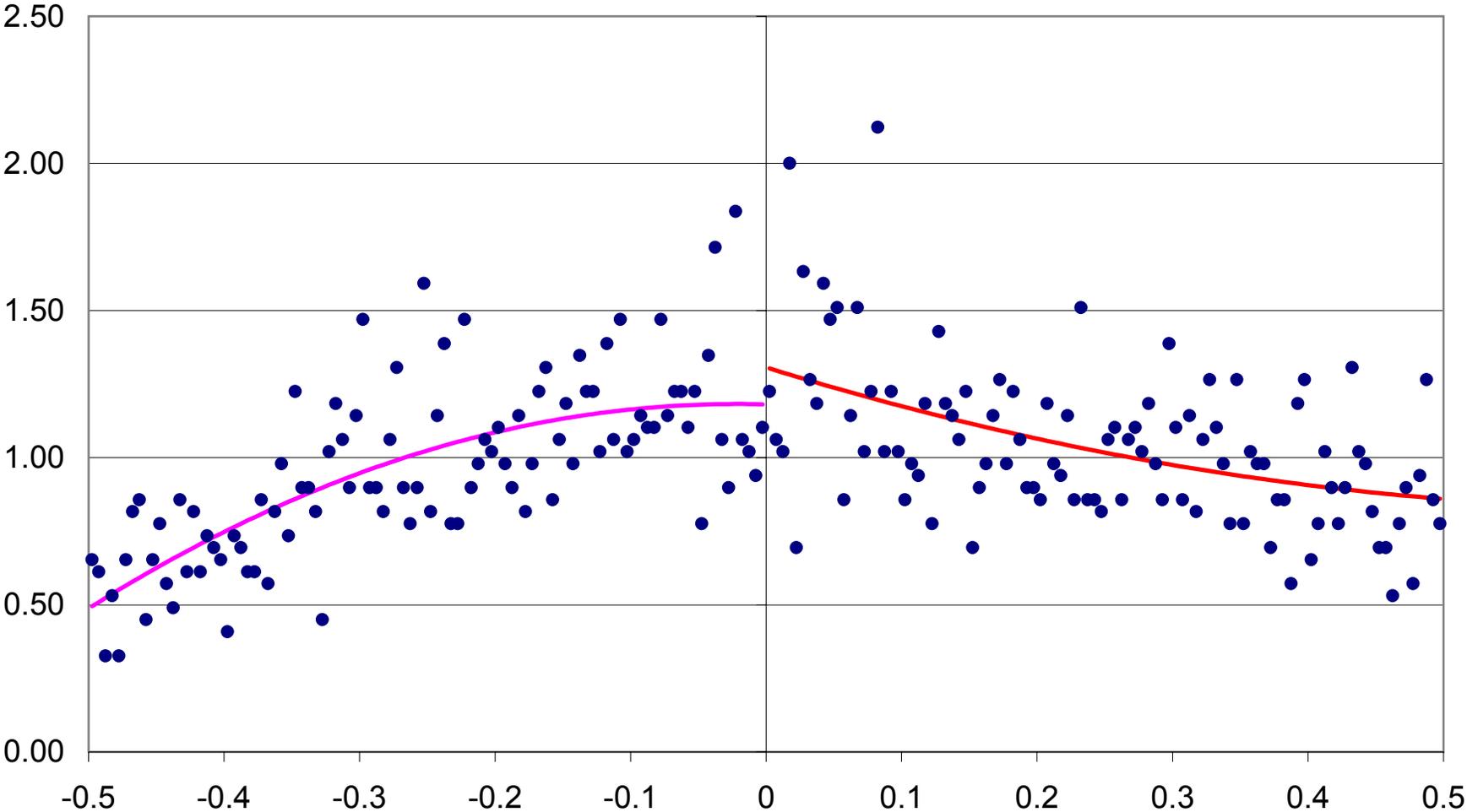


Figure 9: Discontinuity in baseline covariate (share of vote in prior election)

