Methods for Evaluation of Economic Policy Measures

Structural Reforms and Assessment of their Economic Impact

Anže Burger

FDV, UL

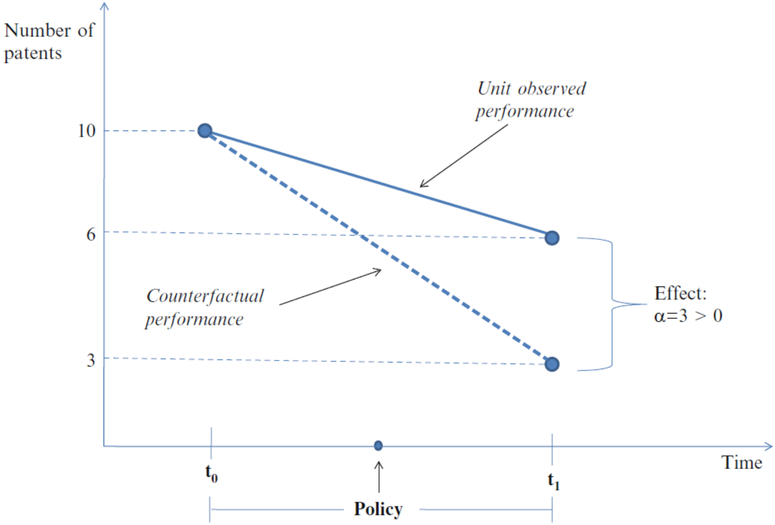
8. marec 2021

Structure of the presentation

1. Introduction to the Econometrics of Program Evaluation
2. Overview of Main Methods Based on Selection on Observables and Unobservables
   - Regression-Adjustment
   - Reweighting
   - Doubly-Robust Estimation
   - Matching
   - Instrumental-Variables Approach
   - Selection-Model
   - Difference-in-Differences
   - Local Average Treatment Effect
   - Regression-Discontinuity-Design
   - Synthetic Control Method

Experimental and quasi-experimental design

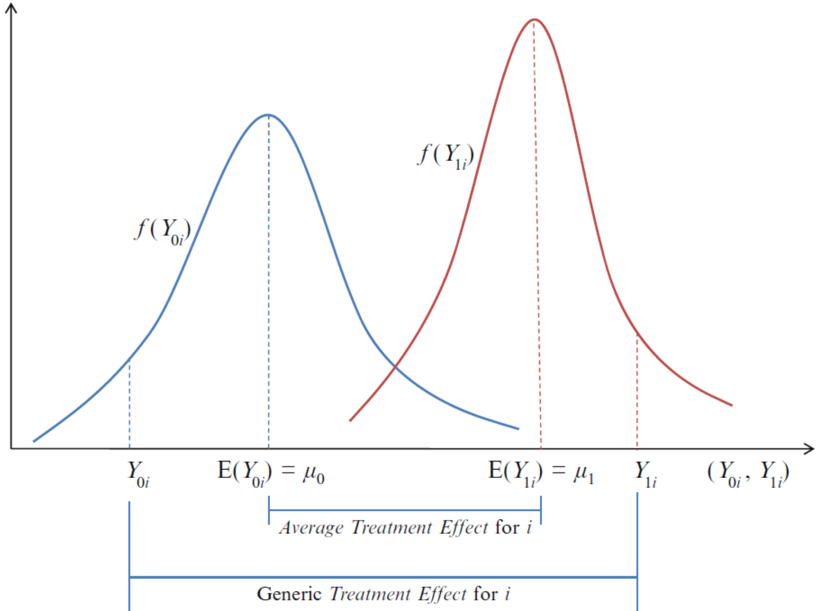
The treatment effect is generally estimated by the counterfactual approach:



Statistical setup

- D is binary treatment indicator (multinomial, continuous)
- Y is outcome variable (binary, continuous, count...)
- Individual i treatment effect: $TE_i = Y_{1i} - Y_{0i}$
- Missing observation problem (Holland 1986)
- Potential Outcome Model: $Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i})$
- Average Treatment Effect: $ATE = E(Y_{1i} - Y_{0i})$
- Average Treatment Effect on the Treated: $ATET = E(Y_{1i} - Y_{0i}|D = 1)$
- Average Treatment Effect on the Non-Treated: $ATENT = E(Y_{1i} - Y_{0i}|D = 0)$
- Stable unit treatment value assumption (SUTVA)

Density distributions of Y_{1i}, Y_{0i}, TE and ATE



Statistical setup

- Relation between ATEs:
  $ATE = ATET \cdot p(D = 1) + ATENT \cdot p(D = 0)$
- Given the knowledge of confounding factors $\mathbf{x}$, we can also define the previous parameters as conditional on $\mathbf{x}$:

$$\mathrm{ATE}(\mathbf{x}) = \mathrm{E}(Y_{1i} - Y_{0i}|\mathbf{x}) \tag{1}$$

$$\mathrm{ATET}(\mathbf{x}) = \mathrm{E}(Y_{1i} - Y_{0i}|D = 1, \mathbf{x}) \tag{2}$$

$$\mathrm{ATENT}(\mathbf{x}) = \mathrm{E}(Y_{1i} - Y_{0i}|D = 0, \mathbf{x}) \tag{3}$$

- From these individual-specific average treatment effects, LIE implies:

$$\mathrm{ATE} = \mathrm{E}_{\mathbf{x}}\{\mathrm{ATE}(\mathbf{x})\} \tag{4}$$

$$\mathrm{ATET} = \mathrm{E}_{\mathbf{x}}\{\mathrm{ATET}(\mathbf{x})\} \tag{5}$$

$$\mathrm{ATENT} = \mathrm{E}_{\mathbf{x}}\{\mathrm{ATENT}(\mathbf{x})\} \tag{6}$$

Identification Under Random Assignment

If the sample is drawn at random (random assignment), it is possible to estimate the ATE as the difference between the sample mean of treated and the sample mean of untreated units, which is the well-known "Difference-in-means" (DIM) estimator.

- independence assumption (IA) holds: $(Y_1; Y_0) \perp D$
- Using POM, we can show DIM is:

$$\begin{aligned} \mathrm{E}(Y \mid D=1) - \mathrm{E}(Y \mid D=0) &= \mathrm{E}(Y_1 \mid D=1) - \mathrm{E}(Y_0 \mid D=0) \\ &= \mathrm{E}(Y_1) - \mathrm{E}(Y_0) = \mathrm{ATE} \end{aligned} \tag{7}$$

- implying also $\mathrm{ATE} = \mathrm{ATET} = \mathrm{ATENT}$

$$\widehat{\mathrm{DIM}} = \frac{1}{N_1} \sum_{i=1}^{N_1} Y_{1,i} - \frac{1}{N_0} \sum_{i=1}^{N_0} Y_{0,i} \tag{8}$$

- The knowledge of **x** is unnecessary for a correct estimation of this casual effect.

Consequences of Nonrandom Assignment and Selection Bias

- Self-selection
- Selection mechanism

When the selection of treated and untreated units is done not randomly, depending on either individual "observable" or "unobservable" characteristics, the DIM estimator is no longer a correct estimation for ATE:

$$\begin{aligned} \mathrm{E}(Y \mid D=1)-\mathrm{E}(Y \mid D=0) &= \mathrm{E}(Y_1 \mid D=1)-\mathrm{E}(Y_0 \mid D=0) \\ &\quad + [\mathrm{E}(Y_0 \mid D=1)-\mathrm{E}(Y_0 \mid D=1)] \\ &= [\mathrm{E}(Y_0 \mid D=1)-\mathrm{E}(Y_0 \mid D=0)] + \mathrm{ATET} \end{aligned}$$

- Selection bias is unobservable since we cannot recover $\mathrm{E}(Y_0 \mid D=1)$.
- Bias vanishes when $\mathrm{E}(Y_0 \mid D) = \mathrm{E}(Y_0)$.

Selection on Observables (Overt Bias)

- Conditional independence assumption (CIA): $(Y_1; Y_0) \perp D|\mathbf{x}$
- For average effects, Conditional mean independence (CMI) suffices: $\mathrm{E}(Y_1|\mathbf{x}, D) = \mathrm{E}(Y_1|\mathbf{x})$ and $\mathrm{E}(Y_0|\mathbf{x}, D) = \mathrm{E}(Y_0|\mathbf{x})$
- From POM and averaging conditional on $(\mathbf{x}, D)$:

$$\begin{aligned} \mathrm{E}(Y \mid \mathbf{x}, D) &= \mathrm{E}\left(Y_0 \mid \mathbf{x}, D\right) + D\left[\mathrm{E}\left(Y_1 \mid \mathbf{x}, D\right) - \mathrm{E}\left(Y_0 \mid \mathbf{x}, D\right)\right] \\ &= \mathrm{E}\left(Y_0 \mid \mathbf{x}\right) + D\left[\mathrm{E}\left(Y_1 \mid \mathbf{x}\right) - \mathrm{E}\left(Y_0 \mid \mathbf{x}\right)\right] \end{aligned} \tag{9}$$

- Under CMI: $ATE(\mathbf{x}) = \mathrm{E}(Y|\mathbf{x}, D = 1) - \mathrm{E}(Y|\mathbf{x}, D = 0) \equiv m_1(\mathbf{x}) - m_0(\mathbf{x}) = m(\mathbf{x})$
- Sample equivalents for ATE and ATET:

$$\widehat{\mathrm{ATE}} = \frac{1}{N}\sum_{i=1}^{N} \widehat{m}(\mathbf{x}_i) \tag{10}$$

$$\widehat{\mathrm{ATET}} = \frac{1}{\sum_{i=1}^{N} D_i}\left[\sum_{i=1}^{N} D_i \widehat{m}(\mathbf{x}_i)\right] \tag{11}$$

Selection on Unobservables (Hidden Bias)

- CI (or CMI) assumption is not sufficient to identify program average effects because:

$$\mathrm{E}(Y_1|\mathbf{x}, D) \neq \mathrm{E}(Y_1|\mathbf{x}) \tag{12}$$

$$\mathrm{E}(Y_0|\mathbf{x}, D) \neq \mathrm{E}(Y_0|\mathbf{x}) \tag{13}$$

- The following bias emerges:

$$\begin{aligned}\mathrm{E}(Y \mid \mathbf{x}, D=1) - \mathrm{E}(Y \mid \mathbf{x}, D=0) =& \mathrm{E}(Y_1 \mid \mathbf{x}, D=1) - \mathrm{E}(Y_0 \mid \mathbf{x}, D=0) \\ & + [\mathrm{E}(Y_0 \mid \mathbf{x}, D=1) - \mathrm{E}(Y_0 \mid \mathbf{x}, D=1)] \\ =& [\mathrm{E}(Y_0 \mid \mathbf{x}, D=1) - \mathrm{E}(Y_0 \mid \mathbf{x}, D=0)] \\ & + \mathrm{ATET}(\mathbf{x})\end{aligned} \tag{14}$$

DIM produces a biased estimation of the causal effect of D on Y that cannot be retrieved observationally as the quantity $\mathrm{E}(Y_0 \mid \mathbf{x}, D=1)$ is unobservable.

Heckman et al. (1998) decomposition of selection bias

As shown above: $DIM = ATET + B_1$, where B_1 is selection bias:
$B_1 = [E(Y_0|D = 1) - E(Y_0|D = 0)]$.
B_1 can be further decomposed into:

$$
\begin{aligned}
\mathrm{B}_1 &= \int_{S_{1\mathbf{x}}} \overline{y}_{01}\, dF(\mathbf{x}, w = 1) - \int_{S_{0\mathbf{x}}} \overline{y}_{00}\, dF(\mathbf{x}, w = 0) \\
&= \underbrace{\int_{S_{1\mathbf{x}} - S_{\mathbf{x}}} \overline{y}_{01}\, dF(\mathbf{x}, w = 1) - \int_{S_{0\mathbf{x}} - S_{\mathbf{x}}} \overline{y}_{00}\, dF(\mathbf{x}, w = 0)}_{\mathrm{B_A} \quad \text{(Bias due to weak overlap)}} +
\end{aligned}
$$

$$
\underbrace{\int_{S_{\mathbf{x}}} \overline{y}_{00}\, [dF(\mathbf{x}, w = 1) - dF(\mathbf{x}, w = 0)]}_{\substack{\mathrm{B_B} \\ \text{(Bias due to weak balancing)}}} + \underbrace{\int_{S_{\mathbf{x}}} \overline{y}_{01}\, dF(\mathbf{x}, w = 1) - \int_{S_{\mathbf{x}}} \overline{y}_{00}\, dF(\mathbf{x}, w = 1)}_{\substack{\mathrm{B_C} \\ \text{(Bias due to selection on unobservables)}}}
$$

Regression-Adjustment

- RA is suitable only when the conditional independence assumption (CIA) holds: $(Y_0; Y_1) \perp D|\mathbf{x}$
- As soon as consistent estimators of $m_1(\mathbf{x})$ and $m_0(\mathbf{x})$ are available, we can estimate causal parameters ATEs through the sample equivalents:

$$\widehat{\text{ATE}} = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i)\right] \tag{15}$$

$$\widehat{\text{ATET}} = \frac{1}{N_1}\sum_{i=1}^{N} D_i \cdot \left[\widehat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i)\right] \tag{16}$$

$$\widehat{\text{ATENT}} = \frac{1}{N_0}\sum_{i=1}^{N} (1 - D_i) \cdot \left[\widehat{m}_1(\mathbf{x}_i) - \widehat{m}_0(\mathbf{x}_i)\right] \tag{17}$$

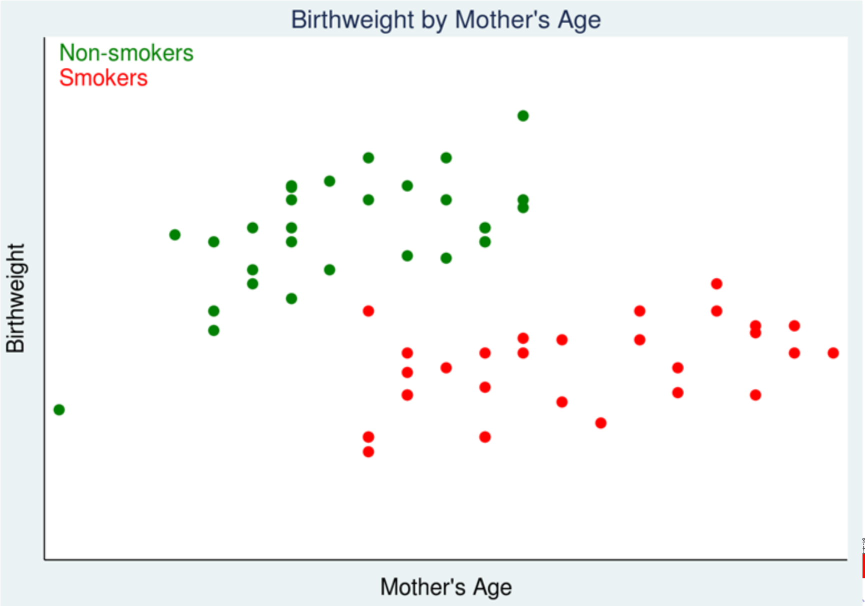
where $m_1(\mathbf{x}) = E(Y|\mathbf{x}, D = 1)$ and $m_0(\mathbf{x}) = E(Y|\mathbf{x}, D = 0)$.

Regression-Adjustment

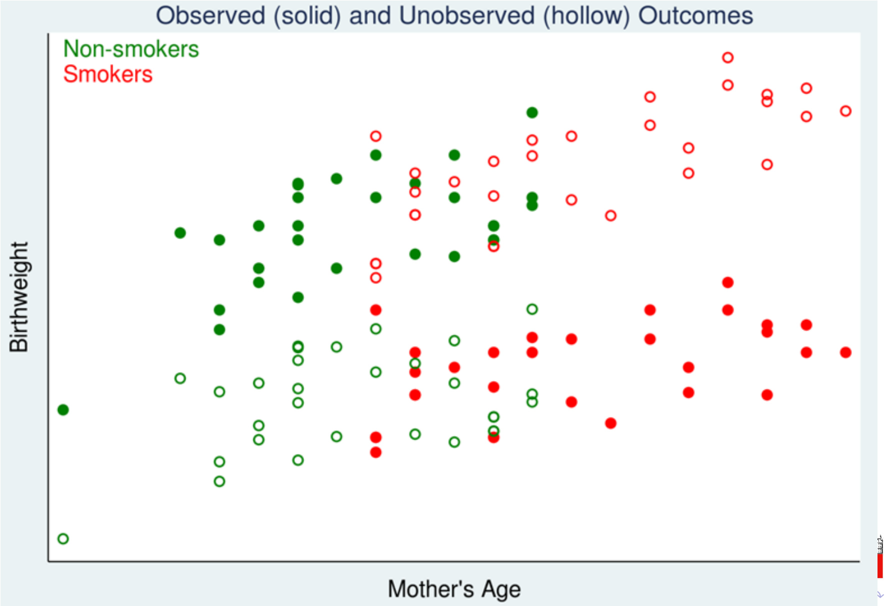
An example explaining the estimation logic of the Regression-adjustment:

Unit	D	x	$m_1 = E(Y \mid D=1;x)$	$m_0 = E(Y \mid D=0;x)$	$m_1 - m_0$	ATET	ATENT	ATE
1	1	A	25	**68**	**–43**			
2	1	B	65	**25**	**40**	**–1.5**		
3	1	C	36	**74**	**–38**			
4	1	D	47	**12**	**35**			
5	0	B	**65**	25	**40**			
6	0	D	**47**	12	**35**			**6.3**
7	0	D	**47**	12	**35**		**11.5**	
8	0	A	**25**	68	**–43**			
9	0	C	**36**	74	**–38**			
10	0	B	**65**	25	**40**			

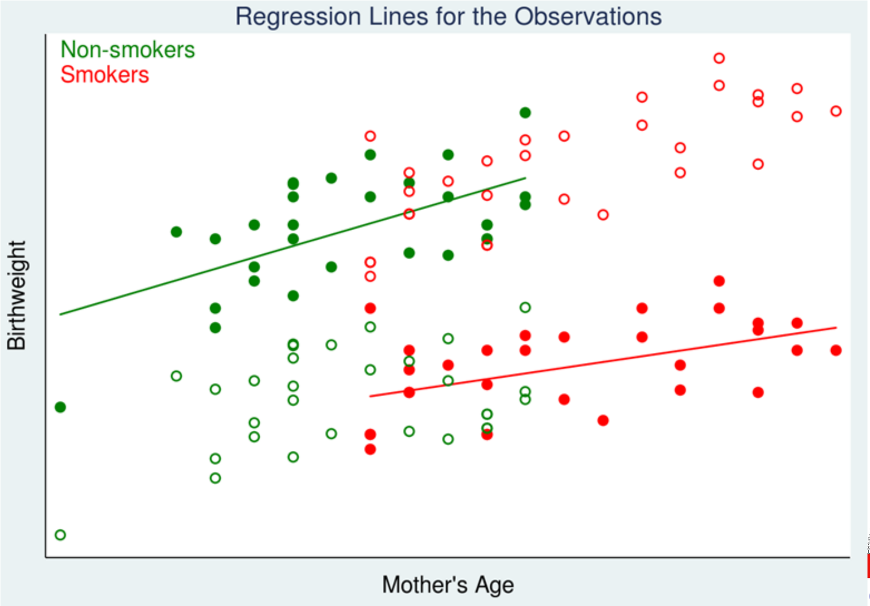
Regression-Adjustment



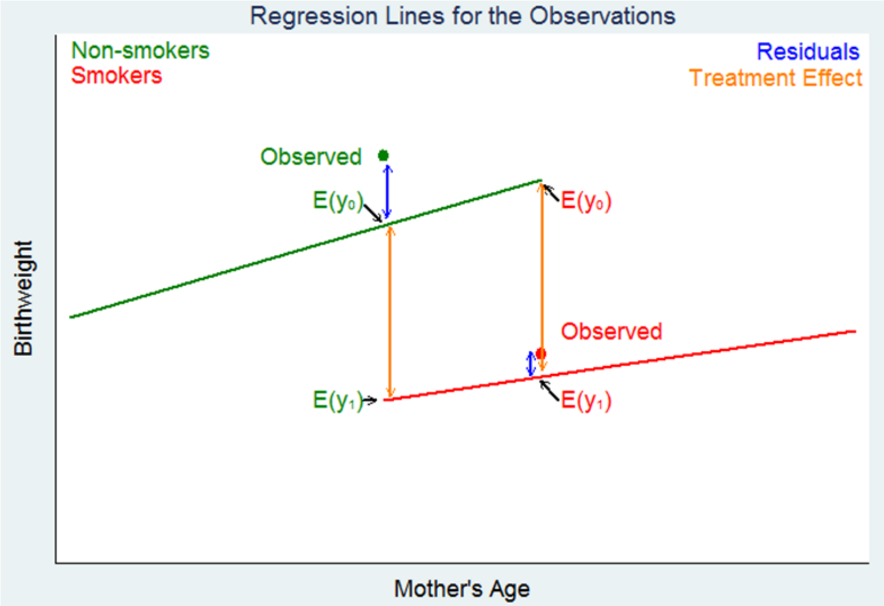
Regression-Adjustment



Regression-Adjustment



Regression-Adjustment



Li et al. (2008): Estimating Average Treatment Effects with Continuous and Discrete Covariates: The Case of Swan-Ganz Catheterization

- Estimate treatment effects for Swan-Ganz catherization (right heart catheterization) for critically ill patients admitted to the ICU.
- $Y_t(w)$ is a binary outcome: it equals 1 if a patient dies within 180 days of being admitted, and 0 otherwise.
- ATE conditional on $X_i = x$: $\tau(x) = \mu_1(x) - \mu_0(x)$ where $\mu_0(x) = E[Y_i(0)|X_i = x]$ and $\mu_1(x) = E[Y_i(1)|X_i = x]$.
- $\tau_{ATE} = E[\tau(X_i)]$ and $\tau_{ATET} = E[\tau(X_i|W_i = 1)$.
- $\widehat{\mu}_w(x)$, $w = 0, 1$ are fitted probabilities using linear regression, MLE logit and kernel estimation methods.
- In addition to RA, they use also propensity score weighting.
- For ICU patients as a whole, applying RHC has no effect on death outcomes.

Regression-Adjustment: Li et al. (2008)

TABLE 1—ESTIMATED TREATMENT EFFECTS OF SWAN-GANZ CATHETERIZATION ON PROBABILITY OF DEATH FOR ICU PATIENTS

	ATE	ATT
Regression	0.076	0.064
(Linear)	(0.050, 0.101)	(0.051, 0.101)
Regression	0.077	0.065
(Logit)	(0.051, 0.101)	(0.051, 0.102)
Regression	0.023	0.022
(Nonparametric)	(0.014, 0.032)	(0.014, 0.032)
Propensity score	0.072	0.063
(Logit)	(0.044, 0.099)	(0.037, 0.089)
Propensity score	−0.0001	0.071
(Nonparametric)	(−0.039, 0.010)	(0.066, 0.118)

Note: Figures in parentheses are the upper and lower bounds of the bootstrapped 95 percent confidence interval.

Reweighting

- When the treatment is not randomly assigned, we expect that the treated and untreated units present very different distributions of their observable characteristics.
- To reestablish some balance in the covariates' distributions, a suitable way could be that of weighting the observations by suitable weights and then using a Weighted least squares (WLS) framework.
- RW do not require one to estimate the regression functions $m_0(\mathbf{x})$ and $m_1(\mathbf{x})$, but they provide estimations of ATEs only by relying on an estimation of $p(\mathbf{x})$, the propensity-score.
- Reweighting approach can be inconsistent either if the specification of the explanatory variables is incorrect or the parametric probit/logit approach does not properly explain the conditional probability of becoming treated.

Reweighting

A general formula for the Reweighting estimator of ATEs takes the following form:

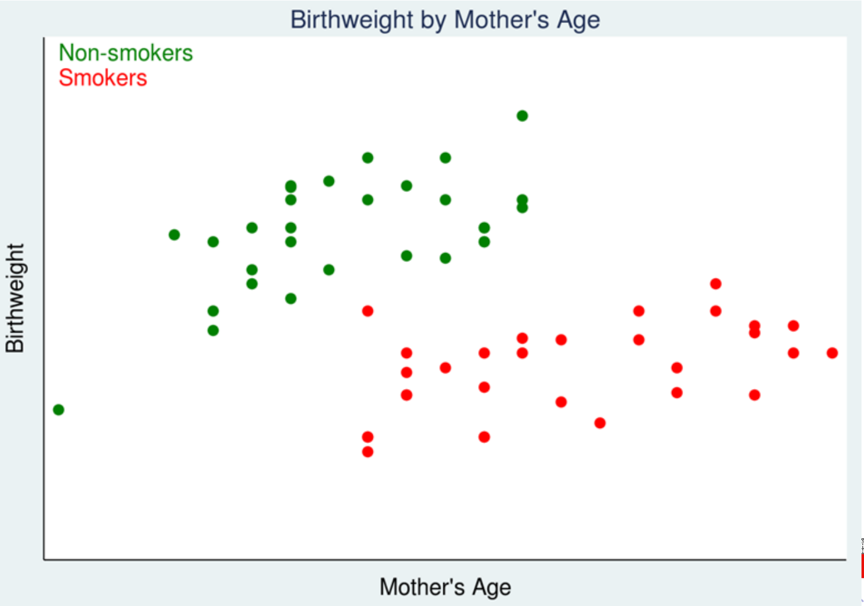
$$\widehat{\mathrm{ATE}} = \frac{1}{N_1}\sum_{i=1}^{N} \omega_1(i) \cdot D_i \cdot Y_i - \frac{1}{N_0}\sum_{j=1}^{N} (1 - D_j) \cdot \omega_0(j) \cdot Y_j \quad (18)$$

$$\widehat{\mathrm{ATET}} = \frac{1}{N_1}\sum_{i=1}^{N} D_i \cdot Y_i - \frac{1}{N_0}\sum_{j=1}^{N} (1 - D_j) \cdot \omega(j) \cdot Y_j \quad (19)$$

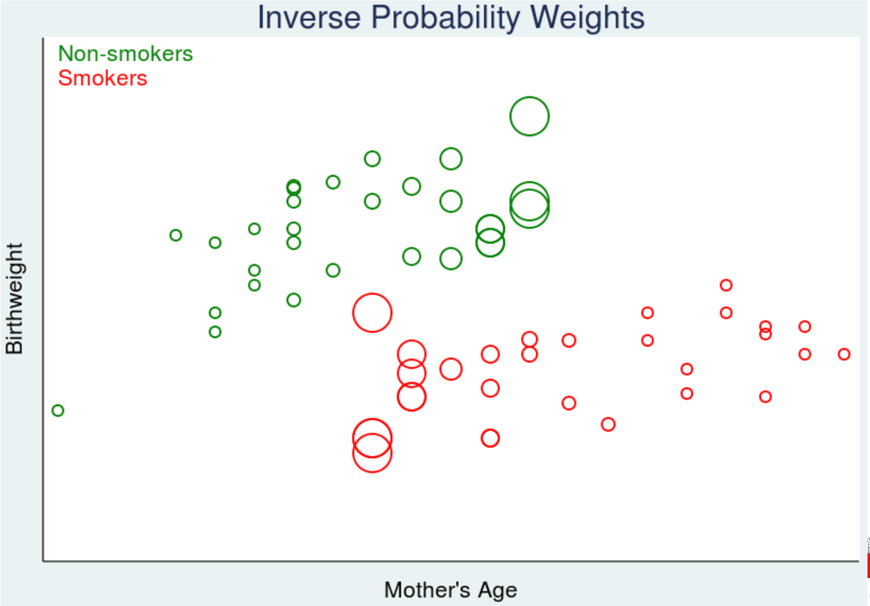
$$\widehat{\mathrm{ATENT}} = \frac{1}{N_0}\left(N \cdot \widehat{\mathrm{ATE}} - N_1 \cdot \widehat{\mathrm{ATET}}\right) \quad (20)$$

The intuitive idea is that of penalizing (advantaging) treated units with higher (lower) probability to be treated and advantaging (penalizing) untreated units with higher (lower) probability to be treated, thus rendering the two groups as similar as possible.

Reweighting



Reweighting



Guadalupe et al.(2012): Innovation and Foreign Ownership

- Show that multinational firms acquire the most productive domestic firms, which, on acquisition, conduct more product and process innovation.
- Probability that a given firm i in industry s is acquired in year t can be estimated using the following linear approximation:

$$F_{it} = \alpha + \beta\varphi_{it-1} + d_t + d_s + \nu_{it} \tag{21}$$

- Estimate of the effect of acquisition on technology using the panel structure of the dataset and including year fixed effects as follows:

$$I_{it} = \alpha + \gamma F_{it-1} + \sum_j \beta^j X^j_{it-2} + d_t + \eta_i + \epsilon_{it} \tag{22}$$

- RW estimator (22) allows them to control not only for selection into being acquired on time-invariant characteristics of firms (just like the equal-weighted fixed effects regression), but also for time-varying characteristics through the propensity score.

Reweighting: Guadalupe et al. (2012)

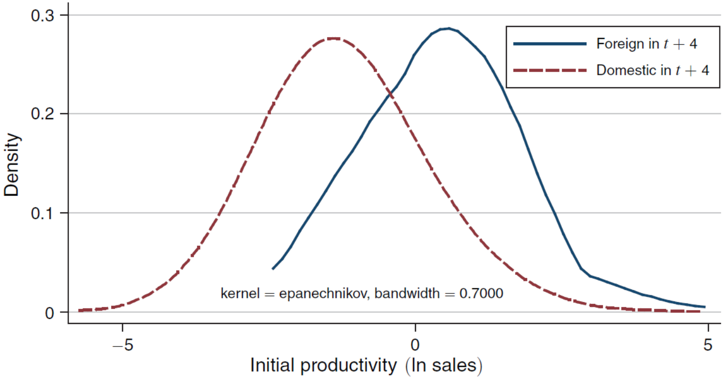


FIGURE 2. DISTRIBUTION OF INITIAL PRODUCTIVITY FOR ACQUIRED AND NONACQUIRED FIRMS

Notes: The dashed line shows the empirical probability density function (pdf) of initial productivity (measured by ln sales demeaned by industry over the sample period) of firms that are domestic at time t and will stay domestic at time $t + 4$. The bold line shows the empirical pdf of initial productivity of firms that are domestic at time t but will become foreign owned by time $t + 4$.

Reweighting: Guadalupe et al. (2012)

TABLE 3—FOREIGN OWNERSHIP AND INNOVATION

	Process innovation				
	(1a)	(2a)	(3a)	(4a)	(5a)
Panel A					
Lag foreign	0.574***	0.419**	0.388*	0.411**	0.611**
	(0.190)	(0.180)	(0.223)	(0.172)	(0.244)
Foreign				0.0459	
				(0.109)	
Forward foreign				0.0663	
				(0.149)	
Observations	20,722	20,671	14,656	12,767	17,578
R^2	0.499	0.527	0.529	0.534	0.532
p-value of test lag foreign = forward foreign				0.0476	
Firm FEs	Yes	Yes	Yes	Yes	Yes
Industry trends		Yes	Yes	Yes	
Selection controls			Yes	Yes	
Propensity score weighting					Yes

Notes: Foreign is an indicator variable that equals one if the firm has at least 50 percent foreign ownership. The dependent variables are our measures of innovation (see Section II for further details). Selection controls include lagged ln firm sales, lagged ln labor productivity, lagged sales growth, lagged export status, lagged average wage, lagged ln capital per employee, lagged ln capital. All columns include year fixed effects. Standard errors are clustered by firm.

Reweighting: Guadalupe et al. (2012)

TABLE 8—FOREIGN OWNERSHIP AND FIRM PRODUCTIVITY

	ln sales					
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
Panel A						
Lag foreign	2.042***	0.165***	0.120**	0.112*	0.0700*	0.182***
	(0.161)	(0.0621)	(0.0599)	(0.0582)	(0.0421)	(0.0540)
Foreign					0.0629	
					(0.0404)	
Forward foreign					−0.0104	
					(0.0646)	
Observations	20,671	20,671	20,671	16,867	14,760	17,578
R^2	0.169	0.100	0.147	0.275	0.284	0.130
p-value of test lag foreign = forward foreign					0.211	
	ln labor productivity					
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
Panel B						
Lag foreign	0.367***	0.126***	0.109**	0.0877	0.109**	0.114**
	(0.0496)	(0.0466)	(0.0449)	(0.0538)	(0.0425)	(0.0487)
Foreign					0.0571	
					(0.0390)	
Forward foreign					−0.0218	
					(0.0425)	
Observations	20,359	20,359	20,359	16,639	14,567	17,338
R^2	0.185	0.014	0.031	0.029	0.035	0.016
p-value of test lag foreign = forward foreign					0.0119	
Firm FEs		Yes	Yes	Yes	Yes	Yes
Industry FEs	Yes					
Industry trends			Yes	Yes	Yes	
Selection controls				Yes	Yes	
Propensity score weighting						Yes

Notes: Foreign is an indicator variable that equals one if the firm has at least 50 percent foreign ownership. ln sales is the natural logarithm of the firm's real sales. ln labor productivity is the natural logarithm of real value added per worker. Selection controls include lagged export status, lagged average wage, lagged log capital per employee, lagged log capital. All columns include year fixed effects. Standard errors are clustered by firm.

Doubly-Robust Estimation

- Combines Reweighting (through an inverse-probability regression) and Regression-adjustment.
- Either the conditional mean or the propensity-score needs to be correctly specified but not both.
- Define a parametric function for the conditional mean of the two potential outcomes as $m_0(\mathbf{x}, \boldsymbol{\delta}_0)$ and $m_1(\mathbf{x}, \boldsymbol{\delta}_1)$, and let $p(\mathbf{x}, \gamma)$ be a parametric model for the propensity-score.
- Estimate $\widehat{p}_i(\mathbf{x}_i)$ by the maximum likelihood (logit or probit).
- Apply a WLS regression using as weights the inverse probabilities to obtain, by assuming a linear form of the conditional mean, the parameters' estimation as:

$$\min_{a_1, \mathbf{b}_1} \sum_{i=1}^{N} D_i \left(y_i - a_1 - \mathbf{b}_1 \mathbf{x}_i\right)^2 / \widehat{p}\left(\mathbf{x}_i\right) \tag{23}$$

$$\min_{a_0, \mathbf{b}_0} \sum_{i=1}^{N} \left(1 - D_i\right) \left(y_i - a_0 - \mathbf{b}_0 \mathbf{x}_i\right)^2 / \left(1 - \widehat{p}\left(\mathbf{x}_i\right)\right) \tag{24}$$

Doubly-Robust Estimation

- Finally, estimate ATEs by Regression-adjustment as:

$$\widehat{\mathrm{ATE}} = 1/N \sum_{i=1}^{N} \left[\left(\widehat{a}_1 - \widehat{\mathbf{b}}_1 \mathbf{x}_i \right) - \left(\widehat{a}_0 - \widehat{\mathbf{b}}_0 \mathbf{x}_i \right) \right] \tag{25}$$

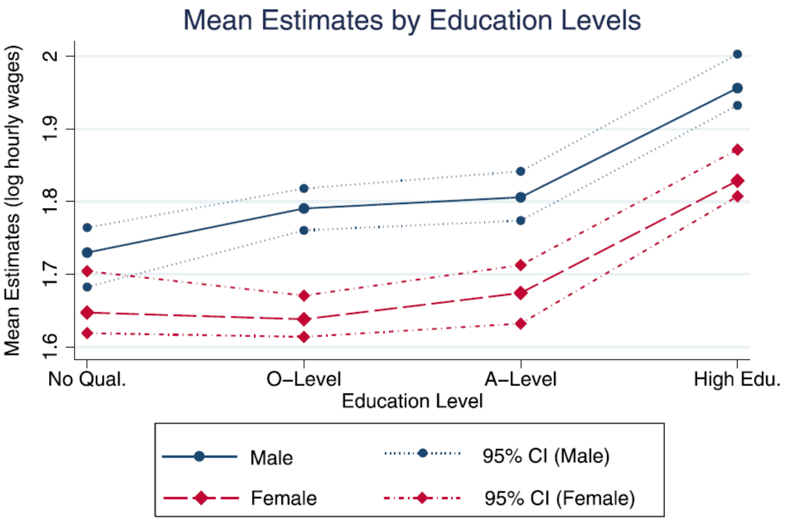
$$\widehat{\mathrm{ATET}} = 1/N_1 \sum_{i=1}^{N} D_i \left[\left(\widehat{a}_1 - \widehat{\mathbf{b}}_1 \mathbf{x}_i \right) - \left(\widehat{a}_0 - \widehat{\mathbf{b}}_0 \mathbf{x}_i \right) \right] \tag{26}$$

$$\widehat{\mathrm{ATENT}} = 1/N_0 \sum_{i=1}^{N} (1 - D_i) \left[\left(\widehat{a}_1 - \widehat{\mathbf{b}}_1 \mathbf{x}_i \right) - \left(\widehat{a}_0 - \widehat{\mathbf{b}}_0 \mathbf{x}_i \right) \right] \tag{27}$$

Uysal (2015): Doubly Robust Estimation of Causal Effects with Multivalued Treatments: An Application to the Returns to Schooling

- Estimate returns to various levels of schooling using unique dataset of the 1970 British Cohort Study with extensive control measures on cognitive and noncognitive abilities, as well as children's behavior.
- Generalize their doubly robust estimator for the potential outcome model with a multivalued treatment variable: No qualification; O-level; A-level; Higher education.
- Relative to no qualification, avg. return to O-levels of 6.3%, to A-levels of 7.9% and to higher educ. of 25.4% for males.
- Average returns to O-level and A-level with respect to no qualification are insignificant for females, whereas the return to higher education is 19.9%.
- Percentage wage gain due to higher education versus O-level and A-level is higher for highly educated females than highly educated males.

Doubly-Robust Estimation: Uysal (2015)



Estimated mean log hourly wages by education level for female and male samp

Matching

Matching is based on recovering the unobservable potential outcome of one unit using the observable outcome of similar units in the opposite status. Its advantages:

- does not require to specify a specific parametric relation between potential outcomes and confounding variables
- wide set of different Matching procedures
- reduces the number of untreated to a subsample of selected controls
- common support

Matching can be performed on:

- covariates (Exact Matching)
- discretized covariates (Coarsened-Exact Matching)
- propensity score (Propensity-Score Matching)

Matching

Matching formulas for ATEs are:

$$\widehat{\mathrm{ATET}}_{\mathrm{M}} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \left[Y_i - \widehat{m}_0 \left(\mathbf{x}_i \right) \right] \tag{28}$$

$$\widehat{\mathrm{ATENT}}_{\mathrm{M}} = \frac{1}{N_0} \sum_{i=1}^{N} \left(1 - D_i \right) \left[\widehat{m}_1 \left(\mathbf{x}_i \right) - Y_i \right] \tag{29}$$

$$\widehat{\mathrm{ATE}}_{\mathrm{M}} = \frac{1}{N} \sum_{i=1}^{N} \left\{ D_i \left[Y_i - \widehat{m}_0 \left(\mathbf{x}_i \right) \right] + \left(1 - D_i \right) \left[\widehat{m}_1 \left(\mathbf{x}_i \right) - Y_i \right] \right\} \tag{30}$$

Matching can eliminate biases B_A (weak overlap) and B_B (weak balancing) but not B_C (unobservable selection) if:

- Conditional mean independence: $E(Y_{1(0)}|\mathbf{x}, D) = E(Y_{1(0)}|\mathbf{x})$
- Overlap: $0 < p(\mathbf{x}) \equiv Pr(D = 1|\mathbf{x}) < 1$

Matching

Table 2.3 Different Matching methods for estimating ATEs according to the specification of C(i) and $h(i, j)$

Matching method	C(i)	$h(i,j)$
One-nearest-neighbor	$\{\text{Singleton } j : \min_j \lVert p_i - p_j \rVert\}$	1
M-nearest-neighbors	$\{\text{First } M\ j : \min_j \lVert p_i - p_j \rVert\}$	$\frac{1}{M}$
Radius	$\{j : \lVert p_i - p_j \rVert < r\}$	$\frac{1}{N_{C(i)}}$
Kernel	All control units (C)	$\frac{K_{ij}}{\sum_{j \in C} K_{ij}}$
Local-linear	All control units (C)	$\frac{K_{ij} L_i^2 - K_{ij} \widehat{\Delta}_{ij} L_i^1}{\sum_{j \in C} \left(K_{ij} L_i^2 - K_{ij} \widehat{\Delta}_{ij} L_i^1 + r_L \right)}$
Ridge	All control units (C)	$\frac{K_{ij}}{\sum_{j \in C} K_{ij}} + \frac{\widetilde{\Delta}_{ij}}{\sum_{j \in C} \left(K_{ij} \widetilde{\Delta}^2_{ij} + r_R h \lvert \widetilde{\Delta}_{ij} \rvert \right)}$
Stratification	All control units (C)	$\frac{\sum_{b=1}^{B} \mathbf{1}[p(\mathbf{x}_i) \in I(b)] \cdot \mathbf{1}[p(\mathbf{x}_j) \in I(b)]}{\sum_{b=1}^{B} \mathbf{1}[p(\mathbf{x}_j) \in I(b)]}$

Bilicka (2019): Comparing UK Tax Returns of Foreign Multinationals to Matched Domestic Firms

- Using confidential UK corporate tax returns data she explores whether there are systematic differences in the amount of taxable profits that multinational and domestic companies report.
- Nearest-neighbour PSM with 0.1 caliper without replacement and Abadie and Imbens (2016) standard errors.
- Foreign multinational subsidiaries underreport their taxable profits by 50 percent relative to domestic standalones.
- Multinational companies are able to use various methods of profit shifting, such as debt shifting, patent or royalty location, or transfer pricing to minimize their taxable profits in the UK. 40 percent of the profit ratio gap can be explained by the differences in leverage.

Matching: Bilicka (2019)

Table 2—Propensity Score Matching: Baseline Results

Sample	Variable	Multinational	Domestic	Diff.	SE	Observations
Baseline	y	0.120	0.238	−0.119	0.011	324,736
Baseline	$y > 0$	0.236	0.285	−0.049	0.020	170,520
Baseline	ztp	0.505	0.205	0.300	0.001	324,736
Positive taxable profits	y	0.234	0.283	−0.049	0.020	170,798

Notes: Results from the baseline propensity score matching estimation, 2000–2014, selected sample. Matching on total assets and within industry and year. Baseline sample estimates unconditional means, positive taxable profits sample estimates means conditional on positive taxable profits, y is the ratio of taxable profits to total assets, ztp is zero taxable profits. Treated observations are foreign multinational subsidiaries, control observations are domestic standalones. Observations column is a sum of total number of observations, which are split equally between the two ownership types.

Sources: Merged HMRC and FAME data

Matching: Bilicka (2019)

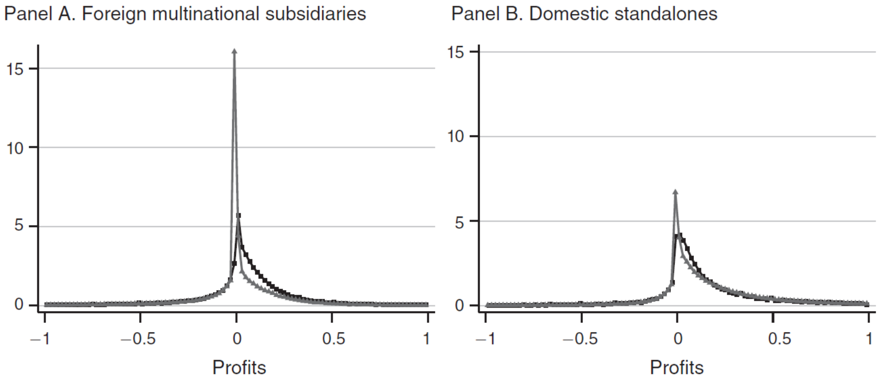


Figure 2. Distributions of Taxable and Accounting Profits: Comparisons

Notes: Distribution of the ratios of taxable profits (including trading losses) from HMRC and profit and loss before taxes from FAME scaled by total assets, propensity score matched sample with non-missing accounting profits data, 2000–2014. Gray line shows distribution of the ratio of taxable profits to total assets, while the black line shows the distribution of the ratio of accounting profits to total assets.

Matching: Bilicka (2019)

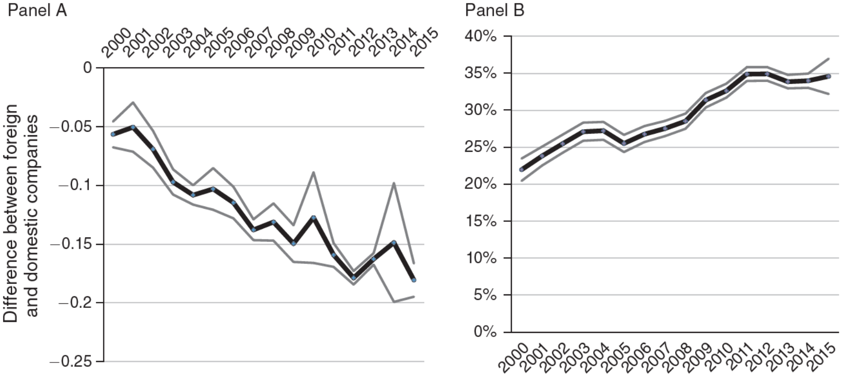


FIGURE 3. PSM: YEARLY HETEROGENEITY

Notes: Results from the Propensity Score Matching estimated year by year. PSM using total assets and within each industry. Panel A: the outcome variable is the ratio of taxable profits to total assets. Panel B: the outcome variable is zero taxable profits dummy. The estimated coefficients for each year are significant. Selected sample, 2000–2014.

Sources: Merged HMRC and FAME data

Matching: Bilicka (2019)

TABLE 5—PSM RESULTS: HEADQUARTER LOCATION HETEROGENEITY

Subsample	Multinational	Domestic	Diff.	SE	Observations
Tax haven	0.096	0.224	−0.128	0.004	55,516
Large tax haven HK SG NL IE	0.097	0.206	−0.110	0.003	63,844
Asian multinationals	0.076	0.177	−0.100	0.004	32,024
French multinationals	0.091	0.176	−0.085	0.005	19,322
US multinational	0.115	0.195	−0.080	0.004	106,122
German multinationals	0.089	0.165	−0.075	0.006	22,334
Other European multinationals	0.128	0.190	−0.062	0.019	41,520
Other foreign multinationals	0.159	0.208	−0.048	0.080	41,590

Notes: Results from the Propensity Score Matching estimates, using total assets and within industry as matching variables. I perform matching for each headquarter subsample to find comparable domestic standalones. The outcome variable is taxable profits/total assets in each row. Observations column is a sum of total number of observations, which are split equally between the two ownership types. Selected sample, 2000–2014.

Sources: Merged HMRC and FAME data

Instrumental-Variables Approach

When selection into a program is driven not only by observables but also by unobservable-to-the-analyst factors, then the CMI no longer holds and RA, (PS)M and RW yield biased ATEs. IV estimation demands at least one instrument z with the following properties (exclusion restriction):

- z is (directly) correlated with treatment D
- z is (directly) uncorrelated with outcome Y

For homogenous case ($u_0 = u_1$): $Y = \mu_0 + D \cdot ATE + \mathbf{x}\beta + u$
Exogeneity of $(\mathbf{x}, z)$ and correlation between D and z give:

$$\begin{aligned} &(a)\ Y_i = \mu_0 + D_i ATE + \mathbf{x}_i\beta + u_i \\ &(b)\ D_i^* = \eta + \mathbf{q}_i\delta + \varepsilon_i \\ &(c)\ D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0 \\ 0 & \text{if } D_i^* < 0 \end{cases} \\ &(d)\ \mathbf{q}_i = (\mathbf{x}_i, z_i) \end{aligned} \tag{31}$$

Instrumental-Variables Approach

In case of zero correlation between u_i and ε_i, OLS is consistent. When $Cov(u_i;\varepsilon_i) \neq 0$ we can rely on 3 methods:

- Direct-2SLS
- Probit-OLS
- Probit-2SLS

For heterogeneous case ($u_0 \neq u_1$):
$Y = \mu_0 + D \cdot ATE + \mathbf{x}\beta_0 + D(\mathbf{x} - \mu_x) + \varepsilon$
where $\varepsilon = e_0 + D(e_1 - e_0)$. There are two subcases in this version:

- $e_1 = e_0$ (only observable heterogeneity)
- $e_1 \neq e_0$ (both observable and unobservable heterogeneities)

Angrist&Krueger (1991): Does Compulsory School Attendance Affect Schooling and Earnings?

- Investigate whether students who attend school longer because of compulsory schooling receive higher earnings as a result of their increased schooling.
- Compulsory schooling laws: require students to remain in school until their sixteenth birthday.
- Individuals born in the beginning of the year start school at an older age, and can therefore drop out after completing less schooling than individuals born near the end of the year.
- The estimated monetary return to an additional year of schooling for those who are compelled to attend school by compulsory schooling laws is about 7.5 percent.

2SLS model:

$$\begin{aligned} E_i &= X_i\pi + \sum_c Y_{ic}\delta_c + \sum_c \sum_j Y_{ic}Q_{ij}\theta_{jc} + \epsilon_i \\ lnW_i &= X_i\beta + \sum_c Y_{ic}\xi_c + \rho E_i + \mu_i \end{aligned} \tag{32}$$

IV Approach: Angrist&Krueger (1991)

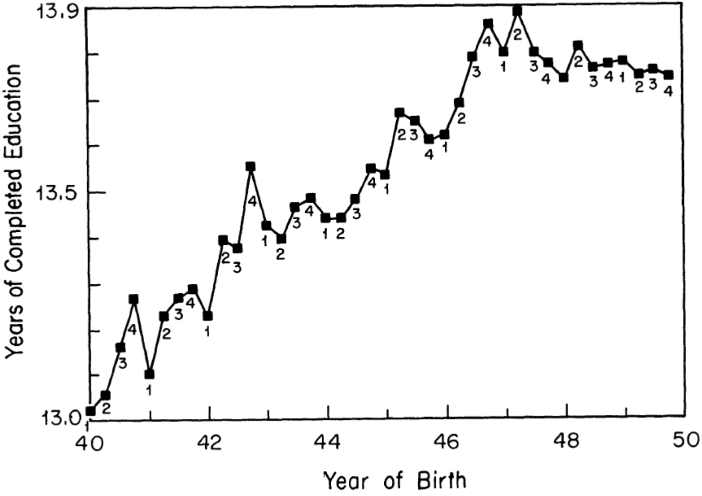


FIGURE II
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

IV Approach: Angrist&Krueger (1991)

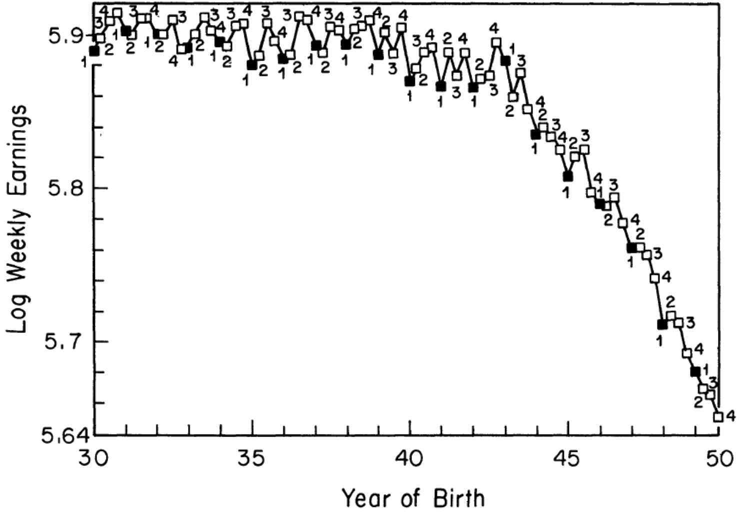


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

IV Approach: Angrist&Krueger (1991)

TABLE IV
OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1920–1929: 1970 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0802	0.0769	0.0802	0.1310	0.0701	0.0669	0.0701	0.1007
	(0.0004)	(0.0150)	(0.0004)	(0.0334)	(0.0004)	(0.0151)	(0.0004)	(0.0334)
Race (1 = black)	—	—	—	—	0.2980	−0.3055	−0.2980	−0.2271
					(0.0043)	(0.0353)	(0.0043)	(0.0776)
SMSA (1 = center city)	—	—	—	—	0.1343	0.1362	0.1343	0.1163
					(0.0026)	(0.0092)	(0.0026)	(0.0198)
Married (1 = married)	—	—	—	—	0.2928	0.2941	0.2928	0.2804
					(0.0037)	(0.0072)	(0.0037)	(0.0141)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region of residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	0.1446	0.1409	—	—	0.1162	0.1170
			(0.0676)	(0.0704)			(0.0652)	(0.0662)
Age-squared	—	—	−0.0015	−0.0014	—	—	−0.0013	−0.0012
			(0.0007)	(0.0008)			(0.0007)	(0.0007)
χ^2 [dof]	—	36.0 [29]	—	25.6 [27]	—	34.2 [29]	—	28.8 [27]

a. Standard errors are in parentheses. Sample size is 247,199. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the State, County, and Neighborhoods 1 percent samples of the 1970 Census (15 percent form). The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

Selection-Model

SM can be compared (if not included) with the IV approach to consistently estimate the parameters in system (31) without the necessity of including an instrument. For lack of instruments, we have to rely on strong assumptions on the joint normality of the error terms.

A generalized Heckit model:

$$
\begin{array}{ll}
(a) & Y = \mu_0 + D \cdot ATE + \mathbf{x}\beta_0 + D(x - \mu_x)\beta + u \\
(b) & E(e_1|\mathbf{x}, z) = E(e_0|\mathbf{x}, z) = 0 \\
(c) & D = [\theta_0 + \mathbf{x}\theta_1 + \theta_2 z + a \geq 0] \\
(d) & E(a|\mathbf{x}, z) = 0 \\
(e) & (a, e_0, e_1) \sim\ {}^3N \\
(f) & a \sim N(0,1) \Rightarrow \sigma_a = 1 \\
(g) & u = e_0 + D(e_1 - e_0)
\end{array}
\tag{33}
$$

Hussinger (2008) R&D and subsidies at the firm level: An application of parametric and semiparametric two-step selection models

- Analyzes the effect of public R&D subsidies on firms' private R&D investment per employee and new product sales in German manufacturing.
- Outcome equation describing the relationship between the $R\&D_i$ intensity, and a vector of covariates X_i:

$$RD_i = X_i' \beta + subsidies_i \delta + \varepsilon_i \tag{34}$$

- Selection equation describes the relationship between a binary participation decision, the receipt of public R&D subsidies, and a vector ofcovariates Z_i:

$$subsidies_i = I Z_i' \gamma + u_i > 0 \tag{35}$$

- The results show that the average treatment effect on the treated firms' R&D intensity is positive.

Selection-Model: Hussinger (2008)

Table III. Estimation of R&D intensity

	OLS	Heckman	Cosslett	Newey	Robinson
Variable	Coef. (t-stat.)				
Subsidies dummy	0.02***	0.04***	−0.02	0.03***	0.02***
	(6.76)	(2.89)	(−0.08)	(5.09)	(5.51)
Log(past subsidies)	−0.10***	−0.11***	−0.12***	−0.10***	−0.10***
	(−8.66)	(−3.59)	(−3.53)	(−3.41)	(−2.91)
Log(past subsidies)^2	0.006***	0.005***	0.006***	0.005***	0.05***
	(8.95)	(3.60)	(3.63)	(3.54)	(3.02)
Log(emp.)	−0.02***	−0.02***	−0.02***	−0.02***	−.02***
	(−7.87)	(−5.96)	(−6.12)	(−5.72)	(−5.72)
Log^2(emp.)	0.001***	0.001***	0.001***	0.001***	0.001***
	(6.12)	(3.87)	(4.71)	(4.36)	(4.20)
Market share	−0.02*	−0.02	−0.01	−0.01	−0.01
	(−1.89)	(−1.18)	(−0.81)	(−0.78)	(−0.34)
Age	−0.02**	−0.02***	−0.02***	−0.02***	−0.02***
	(−2.31)	(−2.79)	(−2.58)	(−2.94)	(−2.87)
Age^2	0.00**	0.00***	0.00***	0.00***	0.00***
	(2.22)	(2.59)	(2.34)	(2.69)	(2.65)
Patent stock	0.01	−0.01	0.01	0.01	0.01
	(0.57)	(−0.13)	(0.12)	(0.06)	(0.09)
Patent stock^2	0.09**	0.12	0.10	0.11	0.11
	(2.31)	(0.83)	(0.75)	(0.69)	(0.70)
Own R&D department	0.01***	0.01***	0.01***	0.01***	0.01***
	(6.42)	(5.74)	(5.53)	(5.60)	(5.09)
Export dummy	−0.002	−0.003	−0.002	−0.002	−0.002
	(−0.84)	(−0.74)	(−0.63)	(−0.62)	(−0.64)
East Germany	0.02***	0.01***	0.02***	0.02***	0.01**
	(6.17)	(2.43)	(4.86)	(5.68)	(1.99)
Constant	0.49***	0.56***	0.57***	0.51***	0.00
	(9.47)	(3.98)	(3.89)	(3.90)	(0.96)
Selection correction					
Mill's ratio funded		−0.01*			
		(−1.82)			
Mill's ratio non-funded		−0.01*			
		(−1.62)			
Newey funded				0.03*	
				(1.63)	
(Newey funded)^2				−0.02*	
				(−1.76)	
(Newey funded)^3				−0.04*	
				(−1.67)	
LR-χ^2 tests					
13 industry dummies	276.87***	277.28***	231.54***	269.55***	169.68***
8 time dummies	52.30***	55.75***	52.07***	50.27***	50.45***
Num. of obs.	3744	3744	3744	3744	3744
F-stat.	32.64	31.01	22.65	30.67	22.37
R^2	0.23	0.22	0.25	0.23	0.16

Difference-in-Differences

DID is suitable in evaluation contexts where observational data for treated and untreated units are available both before and after treatment. Available for repeated cross section and pure panel.

- Two time periods $t = t_0, t_1$; two regimes $s = R, M$; N units i and for each potential outcomes Y_{0ist} and Y_{1ist}
- $ATE(s,t) = E(Y_{1ist} - Y_{0ist}|s,t) = \delta = constant = ATE$
- Common trend assumption: $E(Y_{1ist}|s,t) = \gamma_s + \lambda_t$

$$\begin{aligned} Y_{0ist} &= \gamma_s + \lambda_t + e_{0ist} \\ Y_{1ist} &= \gamma_s + \lambda_t + \delta + e_{1ist} \\ Y_{ist} &= Y_{0ist} + D_{st}(Y_{1ist} - Y_{0ist}) \end{aligned} \tag{36}$$

By simple substitution we obtain: $Y_{ist} = \gamma_s + \lambda_t + D_{st}\delta + e_{ist}$
A simple OLS regression of Y on a regime and time variable and on D_{st} provides a consistent estimation of the $ATE = \delta$.
Specification can be augmented with location-specific trend coefficient and/or covariates $\beta \mathbf{x}_{ist}$.

Difference-in-Differences

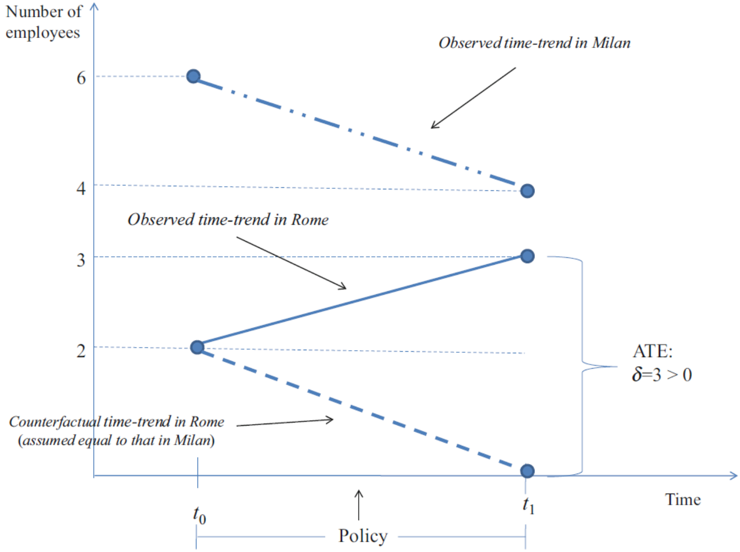


Fig. 3.4 Identification of ATE by the difference-in-differences (DID) estimator

Difference-in-Differences

DID with Panel Data:
DID with panel data can also be easily extended to the case of dynamic treatment by introducing lags and leads:

$$Y_{it} = \gamma_i + \lambda_t + \sum_{\tau=0}^{m} D_{t-\tau}\delta_{-\tau} + \sum_{\tau=1}^{q} D_{t+\tau}\delta_{+\tau} + \beta \mathbf{x}_{it} + e_{it} \tag{37}$$

This can be estimated either by DID (less restrictive identification conditions) or FE regression (more robust).

DID with matching:
A combination of DID with a PSM, it has the advantage that it does not require the imposition of the linear-inparameters form of the outcome equation. In the case of panel data, the M-DID formula takes the following form:

$$\widehat{\mathrm{ATET}}_{\mathrm{M-DID}} = \frac{1}{N_1} \sum_{i \in \{T\}} \left(\left(Y_{i1}^{T} - Y_{i0}^{T} \right) - \sum_{j \in \mathrm{C}(i)} h(i,j) \left(Y_{j1}^{C} - Y_{j0}^{C} \right) \right) \tag{38}$$

Schmeiser et al. (2016): Student Loan Information Provision and Academic Choices

- In 2012 Allen Yarnell Center for Student Success at Montana State University sent warning letters to students with high loan amounts.
- First-semester freshmen with more than \$6,250 in debt, sophomores > \$12,000, juniors>\$18,750, and any student with>\$25,000. Annual tuitions were around \$6,500.
- 57,334 in-state undergraduates from Montana State University and the University of Montana during 2002-2014
- Difference-in-difference-in-differences approach:

$$Y_{it} = \alpha_0 + \beta_1 Letter_{it} + \beta_2 MSU_{it} + \beta_3 Letter_{it} \times MSU_{it} + \beta_4 Letter_{it} \times MSU_{it} \times 2012_{it} + \alpha_1 Demographic_i + \alpha_2 Academic_{it} + \gamma_{semester} + \delta_{year} + \epsilon_{it} \quad (39)$$

Students who receive warning letters are 2% points more likely to switch majors in the semester after receiving the letter,particularly likely into business-related fields and out of health (nursing).

Difference-in-Differences: Schmeiser et al. (2016)

Table 1—Effect of Letters on Student Majors

	Change major	Change into				
		Business	Education	Health	Liberal arts	Science
All students						
β_4	0.020***	0.011*	−0.005	−0.016**	0.004	0.009
	(0.007)	(0.006)	(0.005)	(0.006)	(0.009)	(0.009)
Observations	236,855	236,855	236,855	236,855	236,855	236,855
Low GPA (< 3.0) *students*						
β_4	0.022**	0.010	−0.010*	−0.021***	0.015	0.011
	(0.010)	(0.009)	(0.006)	(0.008)	(0.011)	(0.013)
Observations	110,505	110,505	110,505	110,505	110,505	110,505
High GPA (> 3.0) *students*						
β_4	0.013	0.010	−0.001	−0.016*	−0.013	0.016
	(0.009)	(0.008)	(0.007)	(0.009)	(0.012)	(0.012)
Observations	125,695	125,695	125,695	125,695	125,695	125,695
All freshmen						
β_4	0.032*	0.036***	−0.007	−0.000	−0.038**	0.040*
	(0.017)	(0.013)	(0.008)	(0.013)	(0.019)	(0.022)
Observations	49,163	49,163	49,163	49,163	49,163	49,163
Low GPA (< 3.0) *freshmen*						
β_4	0.021	0.044**	−0.015*	0.006	−0.041	0.011
	(0.025)	(0.019)	(0.009)	(0.018)	(0.025)	(0.029)
Observations	24,913	24,913	24,913	24,913	24,913	24,913
High GPA (> 3.0) *freshmen*						
β_4	0.043*	0.022	0.001	−0.009	−0.042	0.095***
	(0.023)	(0.019)	(0.015)	(0.020)	(0.028)	(0.032)
Observations	24,248	24,248	24,248	24,248	24,248	24,248

Difference-in-Differences: Schmeiser et al. (2016)

TABLE 2—ECONOMIC OUTCOMES BY COLLEGE MAJORS

Field	Unemp rate	Average salary	Default rate
Computer science	7.3	$66,103	1.2
Engineering	3.3	$72,014	1.5
Science, math, ag	5.9	$44,294	5.1
Social science	8.7	$41,316	4.8
Humanities	8.6	$36,197	6.7
Health care	2.0	$52,899	5.8
Business	6.8	$53,126	5.0
Education	6.3	$39,910	6.1

Local Average Treatment Effect

LATE is identified in the setting characterized by randomization under imperfect compliance.

- Let z represent random assignment and D is the actual treatment status and $z \neq D$ (imperfect compliance)
- z is correlated with D but uncorrelated with potential outcome Y
- If treatment effect is heterogeneous over observations this case, it can be proved that the Wald estimator does not consistently estimate the ATE, but LATE (only for compliers)

$$\text{LATE} = \frac{\mathrm{E}(Y \mid z=1) - \mathrm{E}(Y \mid z=0)}{p(D=1 \mid z=1) - p(D=1 \mid z=0)} \tag{40}$$

Consistent estimation of LATE can be obtained from an IV estimation of α in the following regression: $Y = \mu + \alpha D + \mathit{error}$ using z as instrument for D. LATE is equal to the ATET when $p(D_i = 1|z_i = 0) = 0$.

Oreopoulos (2006): Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter

- Legislation from Great Britain's 1944 Education Act raised the school-leaving age in England, Scotland, and Wales in 1947 from 14 to 15 years. Similar in Northern Ireland in 1957.
- Within two years of this policy change, the portion of 14-year-olds leaving school fell from 57 percent to less than 10 percent.
- Raising the school-leaving age to 15 increased earnings, on average, by between 10 and 14 percent.

LATE: Oreopoulos (2006)

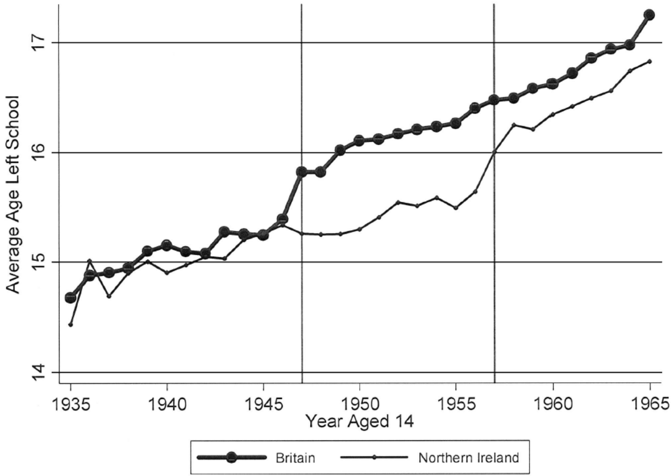


FIGURE 8. AVERAGE AGE LEFT FULL-TIME EDUCATION BY YEAR AGED 14
(*Great Britain and Northern Ireland*)

Note: The upper dark line shows the average age left full-time education by year aged 14 for British-born adults aged 32 to 64 from the 1983 to 1998 General Household Surveys. The lower light line shows the same, but for adults in Northern Ireland.

LATE: Oreopoulos (2006)

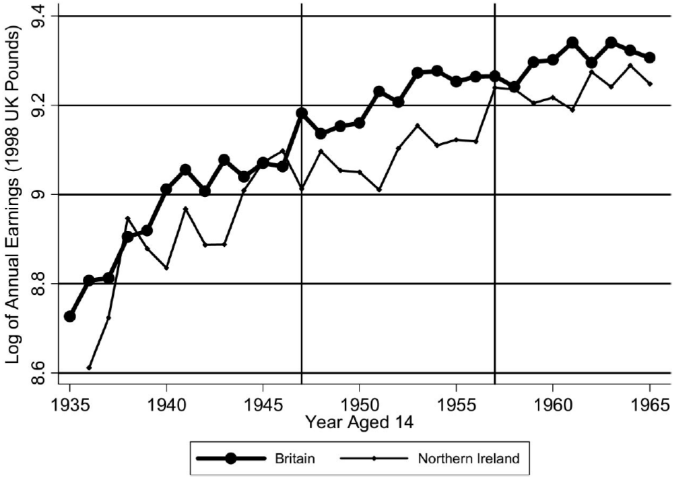


FIGURE 9. AVERAGE LOG ANNUAL EARNINGS BY YEAR AGED 14
(Great Britain and Northern Ireland)

Note: The upper dark line shows the average log annual earnings by year aged 14 for British-born adults aged 32 to 64 from the 1983 to 1998 General Household Surveys. The lower light line shows the same, but for adults in Northern Ireland.

LATE: Oreopoulos (2006)

TABLE 2—OLS AND IV RETURNS TO (COMPULSORY) SCHOOLING ESTIMATES FOR LOG ANNUAL EARNINGS
(Great Britain and Northern Ireland, ages 25–64, 1935–1965)

	(1)	(2)	(3)	(4)	(5)	(6)	(7) Initial sample size
	Returns to schooling: OLS			Returns to compulsory schooling: IV			
Great Britain	0.078	0.079	0.079	0.147	0.145	0.149	57264
	[0.002]***	[0.002]***	[0.002]***	[0.061]**	[0.063]**	[0.064]**	
Northern Ireland	0.111	0.113	0.113	0.135	0.187	0.21	8921
	[0.004]***	[0.004]***	[0.004]***	[0.071]*	[0.070]**	[0.135]	
G. Britain and N. Ireland with	0.082	0.082	0.083	0.174	0.149	0.148	66185
N. Ireland fixed effect	[0.001]***	[0.001]***	[0.001]***	[0.042]***	[0.044]***	[0.046]***	
Birth cohort polynomial controls	Quartic	Quartic	Quartic	Quartic	Quartic	Quartic	
Age polynomial controls	None	Quartic	None	None	Quartic	None	
Age dummies	No	No	Yes	No	No	Yes	

Notes: The dependent variable is log annual earnings. Each regressions includes controls for a birth cohort quartic polynomial and age left full-time education (instrumented by an indicator whether a cohort faced a school leaving age of 15 at age 14 in columns 4 through 6). Columns 2, 3, 5, and 6 also include age controls: a quartic polynomial and fixed effects where indicated. Each regression includes the sample of 25- to 64-year-olds from the 1983 through 1998 General Household Surveys who were aged 14 between 1935 and 1965. Data are first aggregated into cell means and weighted by cell size. Regressions are clustered by birth cohort and region (Britain or N. Ireland).

Regression-Discontinuity-Design

RDD can be used when the selection-into-program (D) is highly determined by the level assumed by a specific "forcing" variable s, defining a threshold $\bar{s}$ separating treated and untreated units.

- Sharp RDD: relation between D and s is deterministic, thus creating a strict "jump" in the probability of receiving the treatment at the threshold
- Fuzzy RDD: relation is stochastic, producing a milder jump

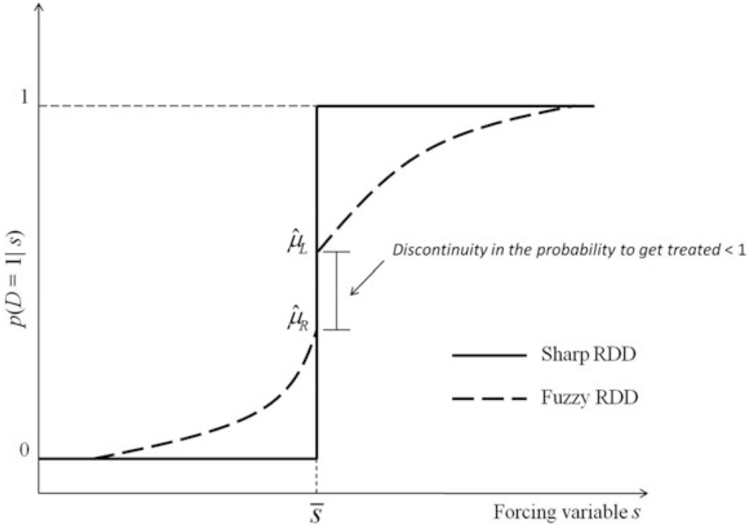
Policy effect is obtained by comparing the mean outcome of individuals laying on the left and the mean outcome of individuals laying on the right of the threshold:

$$\begin{aligned} \mathrm{ATE}_{\mathrm{SRD}} &= \mathrm{E}\left(Y_1 \mid s = \bar{s}\right) - \mathrm{E}\left(Y_0 \mid s = \bar{s}\right) \\ &= \lim_{s \downarrow \bar{s}} \mathrm{E}(Y \mid S = s) - \lim_{s \uparrow \bar{s}} \mathrm{E}(Y \mid S = s) \end{aligned} \tag{41}$$

$$\mathrm{ATE}_{\mathrm{FRD}} = \frac{\lim_{s \downarrow \bar{s}} \mathrm{E}(Y \mid S = s) - \lim_{s \uparrow \bar{s}} \mathrm{E}(Y \mid S = s)}{\lim_{s \downarrow \bar{s}} \mathrm{p}(D = 1 \mid S = s) - \lim_{s \uparrow \bar{s}} \mathrm{p}(D = 1 \mid S = s)} \tag{42}$$

Regression-Discontinuity-Design

Discontinuity in the probability to be treated in the sharp and fuzzy RDD

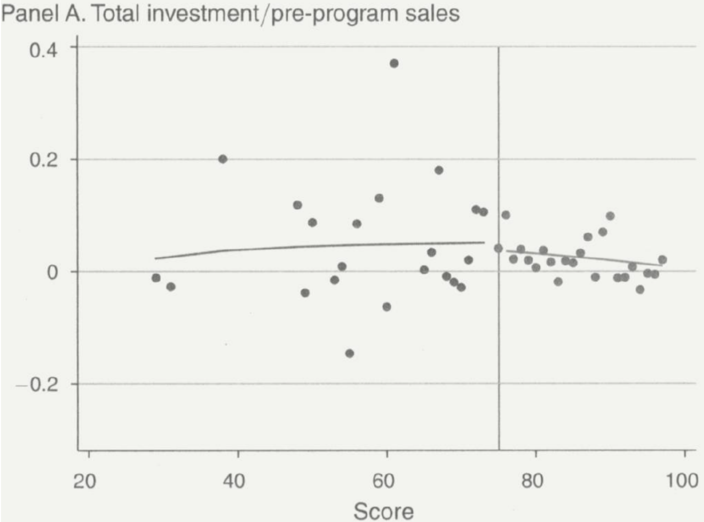


Bronzini&Iachini(2014): Are Incentives for R&D Effective? Evidence from a Regression Discontinuity Approach

- Emilia-Romagna regional government subsidized the R&D expenditure of eligible firms through grants (€93 million, 1246 applicants).
- Only projects deemed sufficient in each category and which obtain a total score of at least 75 points receive the grants (the maximum score is 100).
- Sharp RDD comparing the performance of subsidized and nonsubsidized firms with scores close to the threshold.
- If a subsidy is random around the threshold, treated and untreated firms close to the threshold will be similar.
- Program did not create additional investment: firms substituted public for privately financed R&D.
- SMEs increased their investment substantially, by on average the same amount of the grant received.

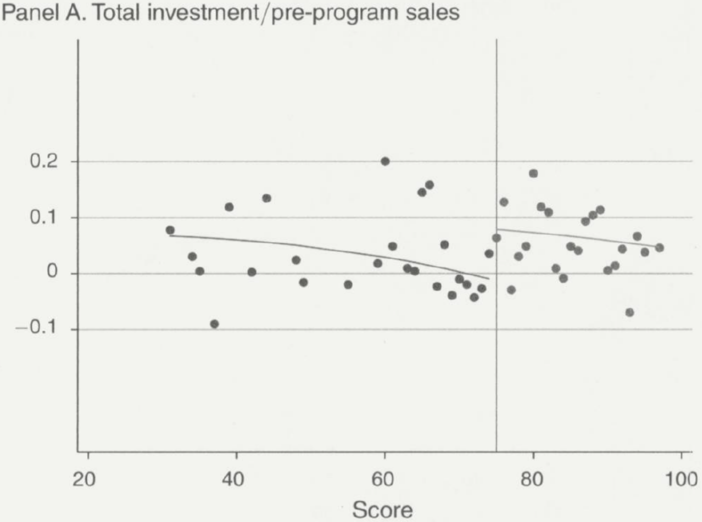
Regression-Discontinuity-Design: Bronzini&Iachini(2014)

Discontinuity in the outcome for large firms



Regression-Discontinuity-Design: Bronzini&Iachini(2014)

Discontinuity in the outcome for small firms



Synthetic Control Method

When the units of observation are a small number of aggregate entities, a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone (i.e. Comparative case studies).

- One treated unit (j=1) and J untreated ("donor pool")
- t=1... $T_0 - 1$ pre-treatment, T_0 intervention and t=$T_0 - 1 \ldots T$ post-intervention periods
- $\mathbf{X}_1$, ..., $\mathbf{X}_{J+1}$ predictors for J units
- Effect of the intervention: $\tau_{1t} = Y_{1t}^I - Y_{1t}^N$

Given a set of weights, $\mathbf{W} = (w_2, ..., w_{J+1})$, the synthetic control estimators of Y_{1t}^N and τ_{1t} are:

$$\widehat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt} \quad (43)$$

$$\widehat{\tau}_{1t} = Y_{1t}^I - Y_{1t}^N \quad (44)$$

Kleven et al. (2013): Taxation and Inter. Migration of Superstars: Evidence from the European Football Market

- Spanish Reform in 2004 (Beckham Law): foreign workers moving to Spain after January 1, 2004 offered a flat tax of 24 percent.
- Danish Reform in 1992 (Tax Scheme for Foreign Researchers and Key Employees): flat tax of 25 percent (30 percent from 1991 to 1995) for a maximum period of 36 months.
- Weights on different countries in the construction of a synthetic control country are nonnegative and chosen to minimize the pre-reform distance between treatment and control in terms of the outcome of interest and indexes of football league quality.
- Clear evidence that international mobility responds to taxation and the effects are stronger for top-quality football players.

Synthetic Control Method: Kleven et al. (2013)

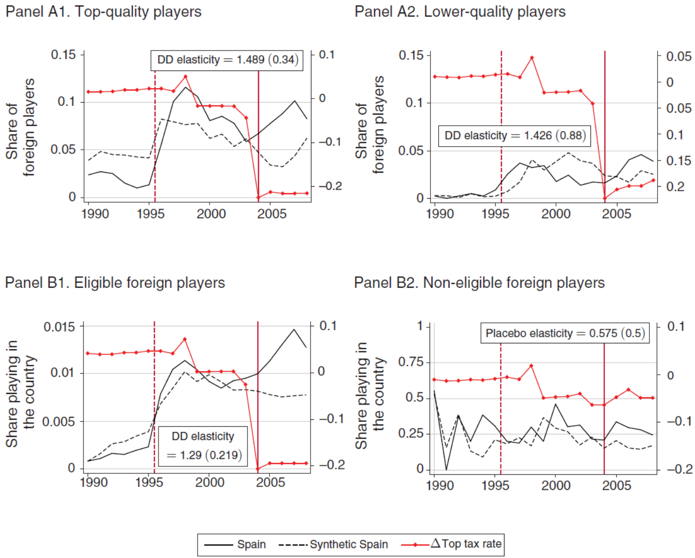


Figure 2. Effects of the 2004 Beckham Law in Spain

Pros&cons of econometric methods for program evaluation

Method	Advantages	Drawbacks
Regression-adjustment (Control-function regression)	*Suitable for observable selection* *Not based on distributional hypotheses*	*Not suitable for unobservable selection* *Based on a parametric estimation*
Matching	*Suitable for observable selection* *Not based on distributional hypotheses* *Based on a nonparametric estimation*	*Not suitable for unobservable selection* *Sensitive to sparseness (weak overlap)* *Sensitive to confounders' unbalancing*
Reweighting	*Suitable for observable selection* *Not based on distributional hypotheses* *Based on a semi-parametric estimation*	*Not suitable for unobservable selection* *Sensitive to propensity-score specification and/or weighting schemes*
Selection-model	*Suitable for both observable and unobservable selection*	*Based on distributional hypotheses* *Based on a parametric estimation*
Instrumental-variables	*Suitable for both observable and unobservable selection* *Not based on distributional hypotheses*	*Availability of instrumental variables* *Based on a parametric estimation*
Regression-discontinuity-design	*Reproducing locally a natural experiment (randomization)* *No distributional hypothesis* *Extendable to nonparametric techniques*	*Availability of a "forcing" variable* *Choice of the cutoff and of an appropriate bandwidth*
Difference-in-differences	*Suitable for both observable and unobservable selection* *Not based on distributional hypotheses*	*Specific form of the error term* *Availability of a longitudinal dataset* *Based on a parametric estimation*

A taxonomy of policy evaluation methods

	Identification assumption		Type of specification		Data structure	
	Selection on observables	Selection on unobservables	Structural	Reduced-form	Cross-section	Longitudinal or repeated cross-section
Regression-adjustment	x			x	x	
Matching	x			x	x	
Reweighting	x				x	
Instrumental-variables	x	x	x		x	
Selection-model	x	x	x		x	
Regression-discontinuity-design	x (sharp)	x (fuzzy)	x (fuzzy)	x (sharp)		
Difference-in-differences	x	x		x		x

Recommended Literature

- Cerulli, G. (2015). *Econometric evaluation of socio-economic programs*. Advanced Studies in Theoretical and Applied Econometrics Series, 49.
- Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10, 465-503.