CONDITIONING ON OBSERVABLES: REGRESSION ADJUSTMENT, WEIGHTING AND MATCHING METHODS

- Reminders
- Identification by CIA
- Methods
- Discussion

Reminders

- We are interested in the impact of a binary treatment $T = (0,1)$
- **Definition:**
  - **Treatment** group: individuals treated ($T = 1$)
  - **Control** group: untreated individuals ($T = 0$)
- For *each* individual *i*, we assume there are two **potential outcomes**:
  - Y_{i0}: potential outcome without treatment = the outcome for individual *i* if she **does not receive** the treatment
  - Y_{i1}: potential outcome with treatment = the outcome for individual *i* if she **receives** the treatment
  - Only one of the two is observed
- We define :

$\Delta_i = Y_{i1} - Y_{i0}$ the causal impact of treatment for individual *i* and

$Y_i = Y_{i0}\,(1 - T_i) + Y_{i1}\,T_i = Y_{i0} + \Delta_i\,T_i$ the observed outcome

Reminders

- **Definition of the parameters of interest**
  - **Mean effect on treatment recipients (ATT or ATET)**

$$\Delta^{ATT} = E(Y_1 - Y_0 | T = 1)$$

    - What impact do benefits have on recipients?
  - **Mean effect on the population (ATE)**

$$\begin{aligned}\Delta^{ATE} &= E(Y_1 - Y_0) \\ &= E(Y_1 - Y_0 | T = 1)P(T = 1) + E(Y_1 - Y_0 | T = 0)P(T = 0)\end{aligned}$$

    - What would the mean effect be if benefits were extended to the entire population? (not always meaningful….)

Reminders

- How can the counterfactual $E(Y_0|T = 1)$ be estimated?
  - The naive estimator $\Delta_0 = E(Y|T = 1) - E(Y|T = 0)$ is **biased** by selection effects…
    - $\Delta_0 = \Delta^{ATT} + B^{ATT}$ where $B^{ATT} = E(Y_0|T = 1) - E(Y_0|T = 0)$
  - ….**unless individuals are randomly assigned** to the treatment and control groups, in which case T is independent of outcome Y_0 and $E(Y_0|T = 0) = E(Y_0|T = 1)$
  - Not generally the case with non-experimental data
    - **Confounding:** treatment and potential outcomes are not independent
    - Information correlated with potential outcomes is used for treatment assignment or there is self-selection based on potential outcomes

- In all what follows, we:
  - focus on **binary treatment**
  - maintain **SUTVA**

Reminders

- **Estimating the causal effect of treatment depends on the <u>assignment mechanism</u>**
  - 'Ideal': **random experiments** (where selection bias are removed as $T_i \perp (Y_{i0}, Y_{i1})$ in this case). *See Rachel Guillain class.*
  - When there are selection effects, at the very least observable differences between beneficiaries and non-beneficiaries must be controlled for:
    - **→ Selection based on observable variables under CIA (or unconfoundness or ignorability) :** $T_i \perp (Y_{i0}, Y_{i1}) \mid X_i$
    - This means having a great deal of information about the members of the population (treated and control)
  - Otherwise for observational (i.e. non experimental) data: differences-in-differences, instrumental variables or regression discontinuity design require specific characteristics

Quiz 1

- You are looking to evaluate the impact of fertility (T_i = 1 if individual *i* has children, 0 otherwise) on women's wages Y_i . Y_{0i} and Y_{1i} are the potential outcomes for individual *i*. What do these terms denote?
  - *Y_0 is the lowest wage she can earn at present, Y_1 the highest.*
  - *Y_0 is the wage she earned before having children, Y_1 her wage since.*
  - *Y_0 is her wage if she has no children at present, Y_1 her wage if she has children.*
  - *Y_0 is her wage if she is not discriminated against due to her fertility, Y_1 her wage if discriminated against.*

Quiz 2

- Which of the following propositions are true of potential outcomes?
  - *Each individual has just one potential income depending on whether they are treated or not.*
  - *There are as many potential outcomes as there are values the treatment can take.*
  - Y_0 *may be higher than* Y_1*.*
  - *Both of an individual's potential outcomes may be observed, for example at different points in time.*

Quiz 3

- Student *X* scores 100 in an English test. In that year student *X* was in a support scheme in English. Students not in the scheme scored 80 in the same test. Student *X* had scored 90 in the same test the year before.
- Without making any assumptions, can we estimate the scheme's impact on student *X*'s score?
  - *Yes, the support scheme meant the student improved by 20 points (Y_1 - Y_0 = 100 - 80 = 20)*
  - *Yes, the support scheme meant the student improved by 10 points (Y_1 - Y_0 = 100 - 90 = 10)*
  - *No, because we do not know Y_1, the score the student would have attained with the scheme.*
  - *No, because we do not know Y_0, the score the student would have obtained without the scheme.*

Identification by CIA

- In the absence of randomization, the identification of treatment effects requires additional assumptions.
- The remainder of this chapter discusses methods based on CIA
  - CIA : confounding, when present is fully accounted for by observed covariates

$$T_i \perp (Y_{i0}, Y_{i1}) \mid X_i$$

  - **In this case, controlling for *X* makes the treatment unconfounded.**
  - This condition, together with the **common support condition** $0 < \Pr(T = 1 \mid X) < 1$ allows identification of treatment parameters as

$$\boldsymbol{E(Y_0|T = 1, X) = E(Y_0|T = 0, X)}$$

  - Unconfoundness is hard to maintain

Identification by CIA

- Large number of methods for estimation and inference under unconfoundedness
  - Regression adjustments
    - Weights are applied to OLS
  - Matching
    - We approximate the results of a random draw by seeking out from the control sample those individuals similar to the treated individuals
    - We look for ‘twins’
  - Inverse probability weighting
  - Hybrid methods

Identification by CIA

- **Limits**
  - CIA is very strong: this means that apart from observable variables, no other characteristics influence both potential outcome and the choice of treatment
    - However, it is often the only method possible → very widely applied
    - Sensitivity analysis
  - For matching, it is not always possible to find an identical untreated individual!
    - **Common support assumption**: for **all** observable values, it must be possible to compare treated and untreated individuals

Methods

- **Method 1: Regression adjustments**
  - Regression: Simplest method
  - *Assumption*: distribution of potential income without treatment conditional on observable characteristics X is linear: $E(Y_{i0}|X_i) = \alpha + \beta X_i$
  - Observed income:

$$y_i = \alpha_0 + \Delta T_i + \beta X_i + \varepsilon_i$$

  - Estimation of the treatment effect Δ by OLS, "controlling" for observables
  - Alternatively, one can use fully saturated models:

$$y_i = \Delta T_i + \Sigma_k \beta_k D_{ki} + \varepsilon_i$$

where D_{ki} is a binary variable taking value 1 if $X_i = x_k$ and 0 otherwise

Methods

- **Note**
  - We can relax the assumption of a constant effect of treatment by adding interaction terms
- **Limits**
  - If the conditional distribution is too remote from the linear form: non-robust estimators
    - *i.e.* varying widely with selected specification
  - The risk is greater when the two samples (control and treatment) have different observable characteristics (Lalonde, 1986)
  - Parametric estimator

Methods

- **Method 2: Matching**
  - We no longer make linear hypothesis
  - We can do matching on covariates or matching on the propensity score
  - In the matching on covariates, each beneficiary is matched with one or more 'twin' non-beneficiaries with exactly the same observables
    - If only one twin is chosen, too much is left to chance
  - It is then necessary to make a further common support assumption : for each beneficiary, there is at least one 'comparable' non-beneficiary

    → the 'nearest neighbour' must be chosen

    → a metric must be defined = measures of 'proximity' in a K-dimensional characteristic space

Methods

- **Estimation**
  - For each beneficiary, we construct an estimator of its counterfactual $\hat{Y}_{i0}$
  - Several methods, the most widely used is the Mahalanobis distance method:

$$d(x_i, x_j) = (x_i - x_j)'\Sigma^{-1}(x_i - x_j)$$

  - The estimated effect is then calculated:

$$\hat{\delta} = \frac{1}{N_1}\sum_{E_1}(Y_{i1} - \hat{Y}_{i0})$$

  - Matching may be done with or without replacement

Methods

- **Problems of the simple nearest neighbour matching method**
  - There is no control over the quality of the matching because the idea of nearest neighbour is relative. However, the method treats near and less-near pairs in the same way
  - Matching with just one individual means missing out on the information provided by all the others, which a priori reduces the precision of the estimation, if some beneficiaries have several very close ‘twins’

Methods

- **Extensions**
  - Extended nearest neighbours: instead of limiting things to the nearest neighbour, we match with a number *M* of nearest neighbours
    - *Radius* (or *caliper*): we select all individuals in a control group located within a fixed neighbourhood
    - *Kernel* (Heckman et al. 1997, 1998): the counterfactual of individual *i* is calculated by kernel estimation. All individuals of the control group are used, but they are weighted by their distance from the treated individual
    - Under certain regularity assumptions, a consistent and asymptotically normal estimator

Methods

- **Problem with these ‘simple’ matching methods**
  - For the independence assumption on observables to be credible: many variables
  - Difficulty in estimating standard deviations
  - **Problem of dimensionality**: properties of convergence and asymptotic normality are not validated for small dimensions (i.e. small number of variables)
  - Non-converging estimation if too many control variables
  - However, on observational data, not credible to think CIA will be validated with few control variables → What can we do?

Methods

- **Propensity score matching (**Rosenbaum and Rubin, 1983):
  - Theoretically overcomes the problem of dimensionality
  - **Definition: propensity score.** Probability of receiving treatment $p(X) = p(T=1|X)$
    - Variables X are the variables explaining participation in programme <u>but not modified by programme</u>
  - **Theorem**: if $Y_0 \perp T \mid X$ (CIA) then income is also independent of treatment conditionally on **propensity score**:

$$Y_0 \perp T \mid X \rightarrow Y_0 \perp T \mid p(X)$$

  - *Interpretation*: same characteristics observable in both groups for values close to $p(X)$

Methods

- **Propensity score estimation**
  - Problem in practice is that this score is unknown and so must be estimated…
  - Two-stage estimation
    - Estimation of score (Logit or Probit)
    - Matching by one of the preceding methods (which is reduced to a one-dimensional problem)
  - Depends on the quality of the first-stage estimate. Check must be made that the score is ‘balanced’, i.e. that observables in both groups are the same for values close to the estimated score
    - There are test to check the score’s ‘balanced’ character

Methods

- **Method 3: Inverse probability weighting**
  - Alternative to matching estimators also based on the propensity score
  - 2 steps
    - First obtain estimates of the propensity score values $\hat{p}(X_i)$
    - Then use those estimates to weight outcome values:

$$\hat{\tau}_{ATE} = \frac{1}{N}\sum_{i=1}^{n}\frac{T_i Y_i}{\hat{p}(X_i)} - \frac{1}{N}\sum_{i=1}^{n}\frac{(1-T_i)Y_i}{1-\hat{p}(X_i)}$$

    Observations with large $\hat{p}(X_i)$ (resp. small) are overrepresented (resp. underrepresented) in the treatment group and thus weighted down (resp. up)

  - Limits: IPW are unstable when the propensity scores approach 0 or 1

Methods

- **Method 4: Hybrid methods**
  - Doubly robust estimators: combination of matching and propensity score weighting with projection and imputation techniques
  - Bias-corrected matching estimators
  - These methods are more stable than IPW

QUIZ 4

- The matching method proposes to evaluate the impact of treatment by comparing the outcome for each treated individual with the outcome for untreated (control) individuals ...
  - *who are similar in terms of observable characteristics*
  - *who are randomly selected*
  - *before and after treatment*

QUIZ 5

- Why is it better to have a large number of observable variables for the matching method?
  - *The identification assumption is more credible.*
  - *Matching is easier.*

QUIZ 6

- What does the propensity score matching method involve?
  - *Matching individuals by their probability of receiving treatment*
  - *Matching individuals by the value of their outcome*
  - *Matching individuals by all observable characteristics*
  - *Matching individuals by the mean of all observable characteristics*

QUIZ 7

- What is (are) the advantage(s) of matching by propensity score?
  - *It reduces bias related to non-observable characteristics that would explain the fact of benefitting from treatment or not.*
  - *It facilitates matching because it is easer to find an untreated individual with a similar propensity score than one with similar observable characteristics.*

QUIZ 8

- It is easier to find for each treated individual an untreated individual with a similar propensity score than an untreated individual with all similar observable characteristics. Even so, this is not systematic: it is possible to find propensity scores in one group for which there are no similar scores in the other group.
- What is the consequence of this?
  - *These observations are removed from the analysis.*
  - *The matching method cannot be used.*
  - *These observations are matched with the closest score even if it is remote.*

QUIZ 9

- Situation: You have to evaluate the impact of a support scheme for high-school pupils in danger of dropping out. The scheme is open to pupils, who decide whether or not to take part. You calculate the propensity score for each pupil, that is, the probability of taking part depending on several observable characteristics: their academic results, age, sex and parents' occupation.
- In the participant group, the probability of receiving treatment is between 50 and 60%. In the non-participant group, the probability of receiving treatment is between 30 and 40%.
- Is the propensity score matching method valid?
  - *No, because participants cannot be matched with non-participants with a similar propensity score.*
  - *Yes, by matching the participants with the highest score with non-participants with the highest score and so on.*

QUIZ 10

- Situation: You have to evaluate the impact of a support scheme for high-school pupils in danger of dropping out. The scheme is open to pupils, who decide whether or not to take part. You calculate the propensity score for each pupil, that is, the probability of taking part depending on several observable characteristics: their academic results, age, sex and parents' occupation.
- Assume now that the propensity score is 40–55% in the treated group and 35–50% in the untreated group.
- Looking more closely at pupils with a propensity score of 40–50%, you see they differ greatly from one group to another. The untreated group is mostly girls from underprivileged backgrounds and the treated group mostly boys from privileged backgrounds.
- Does that reduce the validity of propensity score matching?
  - *Yes, the values of observable characteristics have to be similar*
  - *No, because their propensity scores are similar.*

QUIZ 11

- Situation: You have to evaluate the impact of a support scheme for high-school pupils in danger of dropping out. The scheme is open to pupils, who decide whether or not to take part. You calculate the propensity score for each pupil, that is, the probability of taking part depending on several observable characteristics: their academic results, age, sex and parents' occupation.
- You match participants with non-participants with similar propensity scores. You calculate the difference in outcome on leaving school of each pair, then the mean of that difference to estimate the effect of the scheme.
- Which of the following might bias your analysis?
  - *Pupils in the scheme are older on average than non participants.*
  - *More girls than boys participate in the scheme.*
  - *Pupils in the scheme are driven more by their teachers than non participants.*
  - *Participants in the scheme are more involved in their education than non participants.*
  - *Participants have poorer results than non participants on average.*

QUIZ 12

- Situation: You have to evaluate the effect of a CV (résumé) writing course on the number of interviews job seekers are asked to. The course is open to job seekers on a voluntary basis. You have data on 500 participants and 500 non participants. The data contain observable characteristics (age, education, sex) from which you can construct a propensity score.
- In practice, how can you match job seekers once the propensity score for each of them has been calculated?
  - *You match each participant with a non participant with exactly the same propensity score.*
  - *You match each participant with a non participant with the nearest propensity score.*
  - *You match each participant with several non participants with similar a score.*
  - *You match each participant with all non participants.*

QUIZ 13

- Situation: You have to evaluate the effect of a CV (résumé) writing course on the number of interviews job seekers are asked to. The course is open to job seekers on a voluntary basis. You have data on 500 participants and 500 non participants. The data contain observable characteristics (age, education, sex) from which you can construct a propensity score.
- You choose to match each participant with the non participant with the closest propensity score, rather than with several participants.
- Which of the following cause(s) a particular problem for your choice?
  - *The number of interviews job seekers are asked to varies widely from one month to the next.*
  - *Job seekers who are more involved in their search for employment take the course more.*
  - *The common support zone between the propensity scores of participants and non participants is narrow.*

QUIZ 14

- Situation: You have to evaluate the effect of a CV (résumé) writing course on the number of interviews job seekers are asked to. The course is open to job seekers on a voluntary basis. You have data on 500 participants and 500 non participants. The data contain observable characteristics (age, education, sex) from which you can construct a propensity score.
- You choose to match each participant with several non participants to reduce the impact of variations in the number of interviews job seekers are asked to. How should you weight the number of interviews of each non participant?
  - *You give more weight to non participants with a higher propensity score.*
  - *You give more weight to non participants with a lower propensity score.*
  - *You give more weight to non participants with a propensity score close the participant's.*

QUIZ 15

- Situation: You have to evaluate the effect of a CV (résumé) writing course on the number of interviews job seekers attended in the past month. The course is open to job seekers on a voluntary basis. You have data on 500 participants and 500 non participants. The data contain observable characteristics (age, education, sex) from which you can construct a propensity score.
- Can a participant be matched with a non participant who has already been matched with another participant?
    - *No*
    - *Yes*

Methods

- **Example**
  - Givord et al. (2013) Place-based exemptions and displacement effects: an evaluation of the *Zones Franches Urbaines* program, RSUE
  - Evaluation of the impact of enterprise zones (*ZFU*) on economic activity
    - Tax exemption zones for at least 5 years for firms with less than 50 employees
    - Have they encouraged business and job creation?
    - What effects have they had on neighbouring zones?
  - 'Treated' areas were selected from a pool of underprivileged areas based on a set of characteristics

Methods

- **Example (continued)**
  - Use of panel data for firms (SIRENE and DADS) for the period 2002–2007
  - Data precisely geolocated
  - Variables: date company formed, number of employees, wages and various measures of firm’s economic and financial health (cash flow, debt, investment, etc.)
  - Procedure:
    - 1997: 44 zones become ZFUs and 416 zones become ZRUs
    - 2004: 41 new ZFUs, selected from the ZRU pool based on a composite index combining five indicators
    - Comparison of outcomes between ZRUs that become ZFUs and ZRUs whose status is unchanged
    - Selection bias problem: firms can continue not to move to ZFUs (negative selection bias)
    - Propensity score calculated on basis of logit and matching based on Gaussian kernel

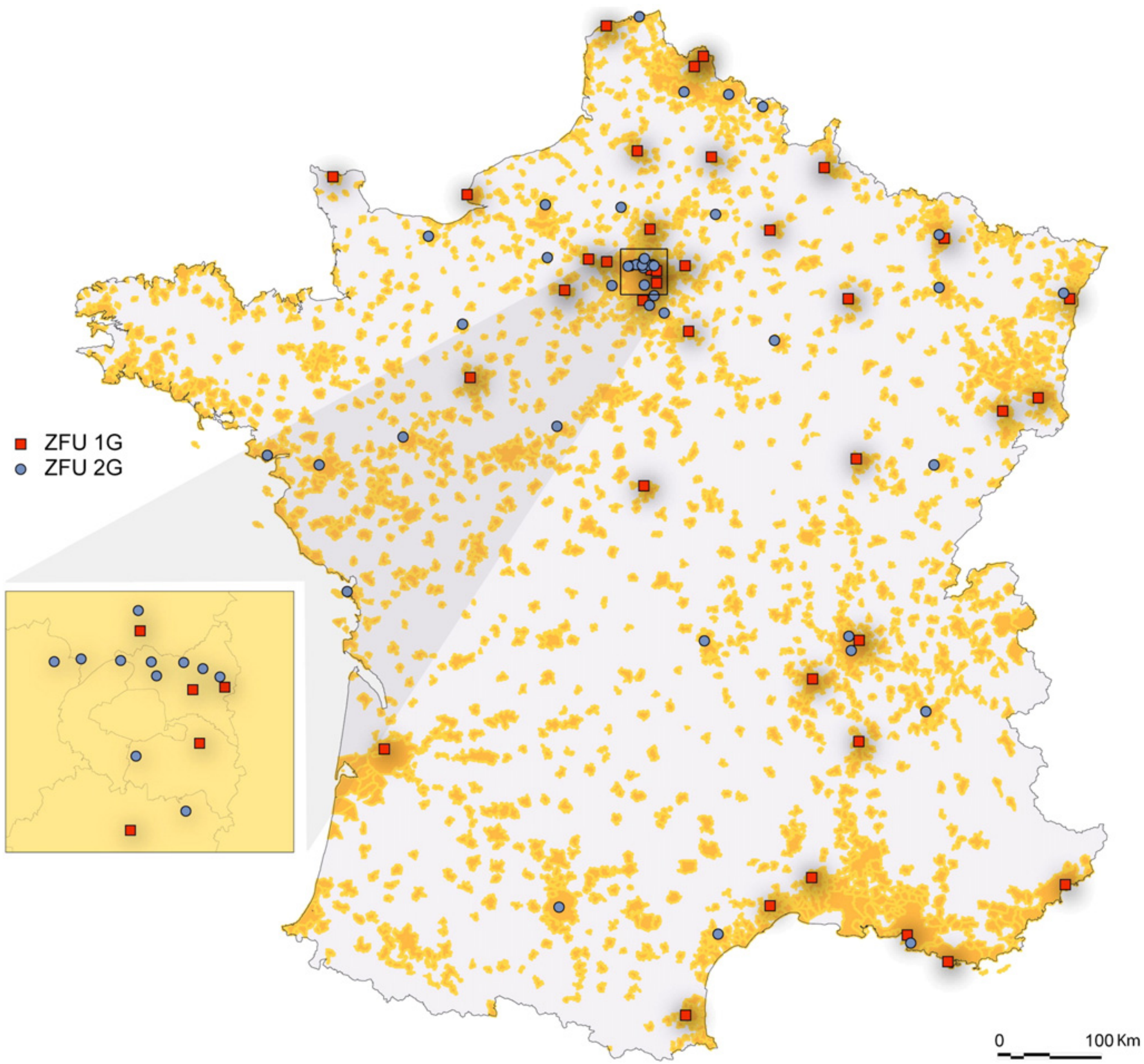


Fig. 1. Distribution of ZFUs created in 1997 (ZFU 1G) and 2004 (ZFU 2G) over the French metropolitan territory. The inset map represents the *Île-de-France* region.

Table 1
Economic situation in the areas in 2002.

	ZRUs not turning ZFU after 2004	ZRUs turning into ZFU in 2004
Total population	5433	12,644
Area (km^2)	77.2	135.4
Number of businesses	84	187
Number of workers	389	685
Business demography (per yr.)		
New businesses	15.27	37.45
Births	12.68	31.47
Transfers	2.59	5.98
New businesses (/ Stock in 2001)	0.19	0.21
Births (/ Stock in 2001)	0.16	0.18
Transfers (/ Stock in 2001)	0.03	0.03
Industries		
Share of manuf.	0.11	0.12
Share of retail	0.36	0.32
Share of construction	0.24	0.25
Share of services to hh.	0.15	0.15
Share of services to b.	0.08	0.10
Share of transportation	0.04	0.04
Number of ZRUs	284	51

Notes: Restriction to eligible companies (50 employees or/and a total revenue above €10 million).

Table 2
Descriptive statistics for eligible companies in 2002.

	Area not turning into ZFU after 2004	Area turning into ZFU in 2002
Number of workers	4.96	5.10
Hourly wage (euro)	10.54	10.64
Sales (×1000 euro)	250.22	232.00
Income (×1000 euro)	35.50	33.03
Cash flow (×1000 euro)	26.61	25.87
Investment (×1000 euro)	12.55	11.58
Debt (×1000 euro)	93.67	93.38
Average number of eligible companies per area	21	38

Table 3
Probability for an area in 2003 to be part of a ZFU in 2004.

Variables	Probability of belonging to a ZFU in 2004
Intercept	496.69
Distance to closest ZRU<2 km	1.35***
Distance to closest ZRU<5 km	1.38**
Distance to closest ZFU1997>30 km	2.42***
Log total population	−165.05
Log total population, squared	18.02
Log total population, cubed	−0.63
Fiscal potential	−1.06
Proportion of less than 25	5.03**
Proportion of dropouts	0.25
Unemployment rate	−0.72
Log number of establishments	−1.10
Δ Log number of establishments	6.57**
Log employment	0.36
Δ Log employment	−0.67
Δ Log employment (present in 2002)	2.14
Δ Log hours worked (present in 2002)	−1.10
(r)1-2 Number of ZFU 2004 observations	51
Number of non-ZFU 2004 observations	284
Efron's R-squared	0.47

Note: Logit estimation. The standard deviation of the estimator is in brackets. Three (respectively two, one) stars indicate a 1% significance (respectively 5%, 10%).

Table 4
Impact of the transition to ZFU on stock of companies and business demography.

Variables	Years				
	2003	2004	2005	2006	2007
Stock (Δ Log)					
Number of establishments	0.01	0.05***	0.07***	0.06**	0.05**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Amongst companies eligible	−0.00	0.05	0.01	−0.04	−0.03
already present in 2002	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
…with less than	−0.04	0.10	0.07	0.02	0.02
3 employees in 2002	(0.04)	(0.07)	(0.06)	(0.05)	(0.08)
…with more than	0.04	0.03	−0.05	−0.06	−0.04
4 employees in 2002	(0.04)	(0.04)	(0.05)	(0.06)	(0.07)
Amongst companies eligible already present in 2002					
Δ Bankruptcies	0.21	0.30	1.28	−2.45	−0.28
(for 1000 companies)	(1.77)	(2.27)	(2.33)	(1.81)	(1.32)
Flow (relatively to the previous stock)					
Δ Births and transfers	0.00	0.06***	0.06*	0.08***	0.05**
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Δ Births	0.01	0.04***	0.04*	0.05***	0.01
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Δ Transfers	−0.00	0.02***	0.02*	0.03***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Note: All results featured herewith correspond to the preferred specification of propensity-score matching, applied to time differentiated variables. The standard deviation of the estimator is in brackets, estimated by block bootstraps in areas. Three (respectively two, one) stars indicate a 1% significance (respectively 5%, 10%).

Table 5
Impact of the transition to ZFU on employment (log).

Variables	Years				
	2003	2004	2005	2006	2007
All companies					
Δ Employment	−0.01	0.04	0.06	0.04	0.12**
	(0.05)	(0.04)	(0.05)	(0.08)	(0.06)
Δ Hours	−0.01	0.03	0.08	−0.00	0.12**
	(0.04)	(0.04)	(0.05)	(0.08)	(0.06)
Amongst companies eligible already present in 2002					
Δ Employment	−0.02	−0.03	−0.06	0.09	0.00
	(0.04)	(0.05)	(0.05)	(0.08)	(0.06)
Δ Hours	0.00	−0.05	−0.07	0.09	−0.02
	(0.04)	(0.05)	(0.05)	(0.08)	(0.08)
…with less than 3 employees in 2002					
Δ Employment	−0.00	−0.08	0.06	0.09	0.12*
	(0.06)	(0.07)	(0.09)	(0.08)	(0.06)
Δ Hours	0.02	−0.10*	0.03	0.14***	0.02
	(0.06)	(0.05)	(0.07)	(0.05)	(0.06)
…with more than 4 employees in 2002					
Δ Employment	−0.05	0.01	−0.05	0.09	0.04
	(0.04)	(0.05)	(0.06)	(0.08)	(0.04)
Δ Hours	−0.03	−0.00	−0.08	0.07	0.07
	(0.04)	(0.05)	(0.07)	(0.09)	(0.05)

Note: All results featured herewith correspond to the preferred specification of propensity-score matching, applied to time differentiated variables. The standard deviation of the estimator is in brackets, estimated by block bootstraps in areas. Three (respectively two, one) stars indicate a 1% significance (respectively 5%, 10%).

Table 6
Impact of the transition to ZFU on wages and accounting indicators, for eligible companies existing in 2002.

Variables	Years				
	2003	2004	2005	2006	2007
All eligible companies present in 2002					
Income (Δ Log)	−0.00	−0.02**	−0.01	−0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sales (Δ Log)	0.00	−0.01	−0.12**	0.06	0.02
	(0.03)	(0.04)	(0.06)	(0.07)	(0.06)
Hourly wage (Δ Log)	−0.00	0.01	−0.00	0.00	0.02
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)
Δ Cash flow/Sales	−0.01	0.01	0.00	−0.03*	0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Δ Debt/Sales	−0.02	−0.04	−0.00	−0.01	0.00
	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)
Δ Investment/Sales	−0.01	0.01	0.01	−0.01	−0.01
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)

Note: All results featured herewith correspond to the preferred specification of propensity-score matching, applied to time differentiated variables. The standard deviation of the estimator is in brackets, estimated by block bootstraps in areas. Three (respectively two, one) stars indicate a 1% significance (respectively 5%, 10%).

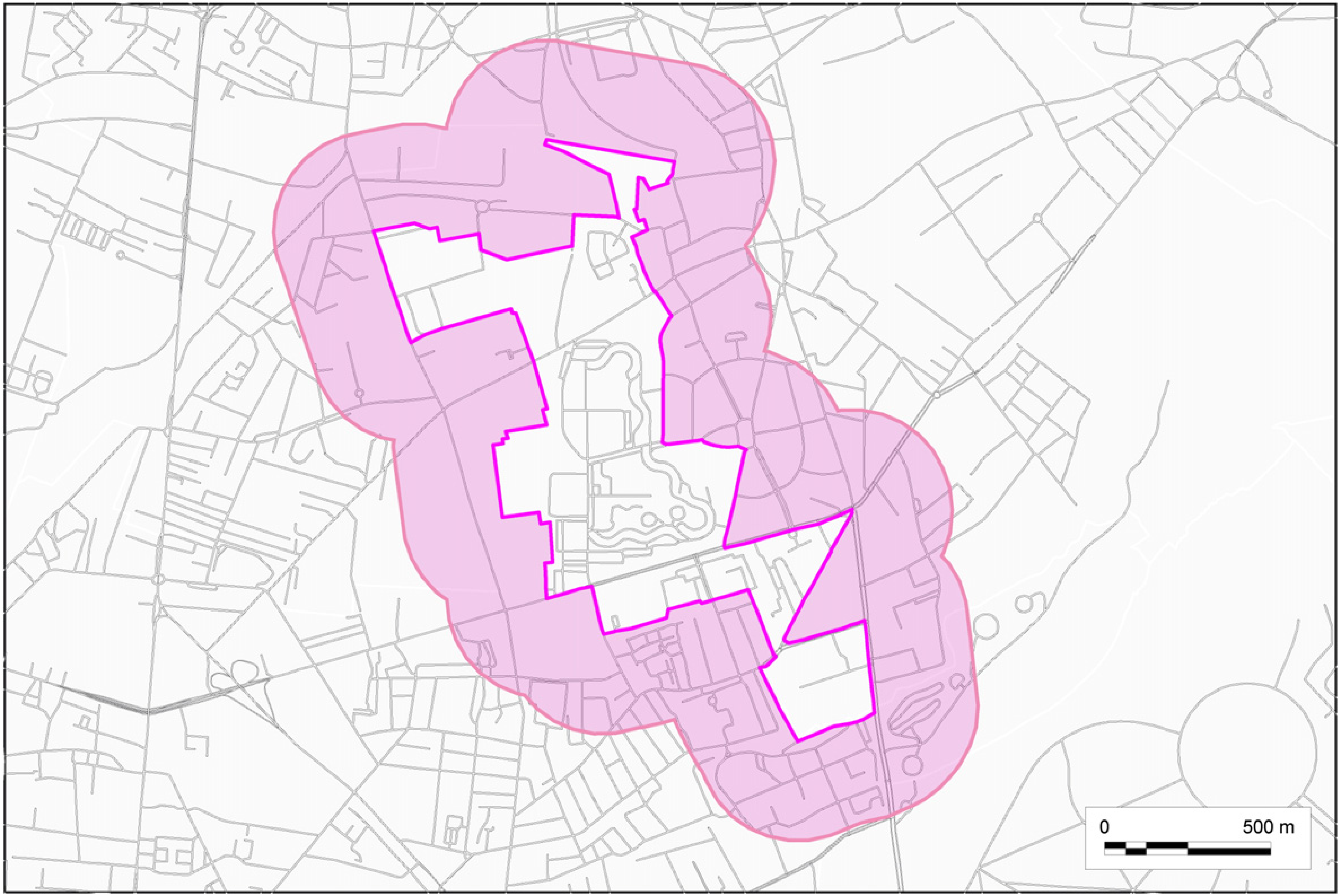


Fig. 9. The ZFU and its 300-meter-wide ring in Stains, in the Northern banlieue of Paris.

Table 8
Impact of the ZFU program on a ring around the treated area.

Variables	Years				
	2003	2004	2005	2006	2007
Number of establishments	0.01	−0.05**	−0.02	−0.09***	−0.08*
	(0.03)	(0.02)	(0.02)	(0.04)	(0.05)
Δ Employment	0.03	−0.07	0.02	−0.08	−0.04
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)
Δ Hours	0.05	−0.10*	0.01	−0.05	−0.07
	(0.06)	(0.05)	(0.05)	(0.07)	(0.10)
Flow (relatively to the previous stock)					
Δ New establishments	−0.05	−0.05	−0.03*	−0.07**	−0.17
	(0.06)	(0.04)	(0.02)	(0.03)	(0.13)
Δ Births	−0.06	−0.04	−0.01	−0.01	−0.06
	(0.06)	(0.04)	(0.02)	(0.02)	(0.04)
Δ Transfers	0.01	−0.01	−0.02*	−0.06***	−0.12
	(0.00)	(0.01)	(0.01)	(0.02)	(0.09)

Note: All results featured herewith correspond to a Gaussian kernel matching method, applied to time differentiated variables. The standard deviation of the estimator is in brackets, estimated by block bootstraps in areas. Three (respectively two, one) stars indicate a 1% significance (respectively 5%, 10%).

Discussion

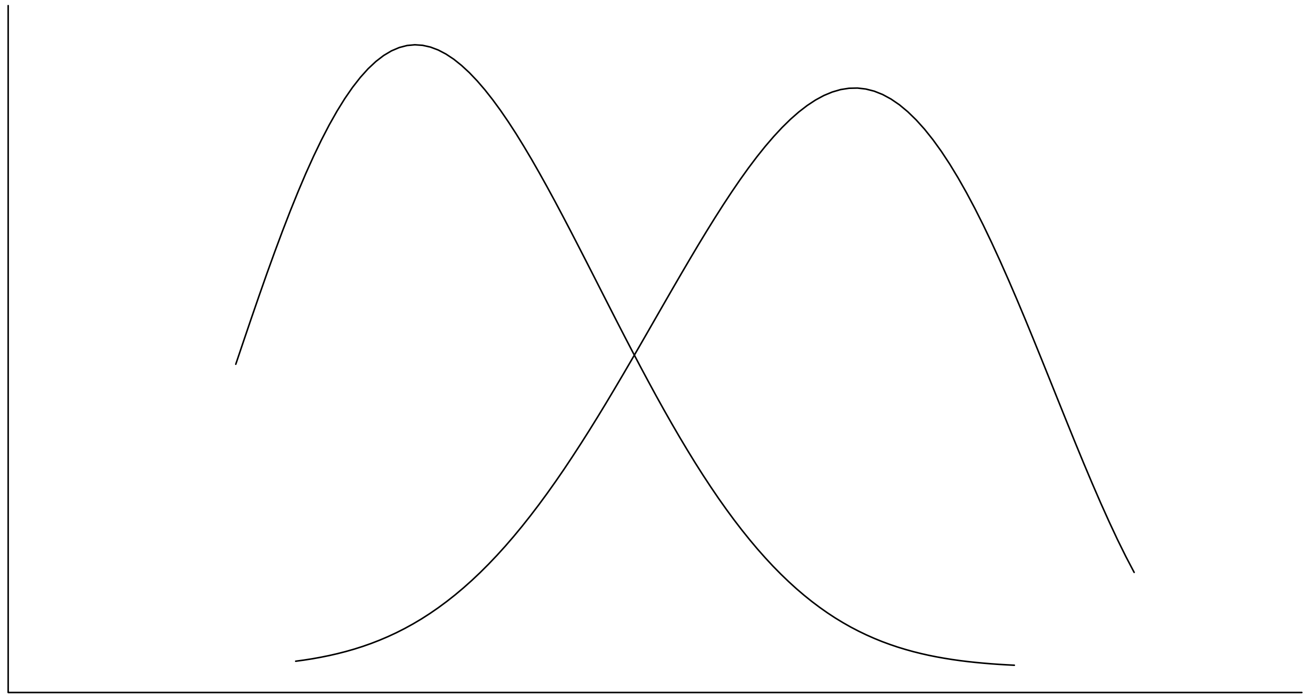
- **Question of inference**
  - Determining the statistical precision of estimators is not straightforward in matching procedures
  - Not always explicit forms for precision, or they are complex forms
  - Precision is estimated by bootstrapping:
    - Sampling with replacement from the sample of N of a new sample of size N: a new estimator can be calculated by the chosen method
    - Repeating this operation K times gives an empirical distribution of the estimator obtained (so standard deviations, confidence intervals, etc. can be estimated)
  - Very flexible method but unsuitable for nearest neighbour matching methods

Discussion

- **Discussion of the common support assumption**
  - The method applies only to the common support of the score distribution for beneficiaries and non beneficiaries
  - It must be checked that the common support overlaps enough
    - Comparison of score distribution on two subsamples
  - Unless we keep to the common support region, estimates might be biased
  - Several methods of keeping to the common support region
  - CAUTION: this changes the nature of what is estimated → local estimator for individuals who are in common support region
    - If common support is restricted, we estimate the impact for individuals in the tail of the distribution

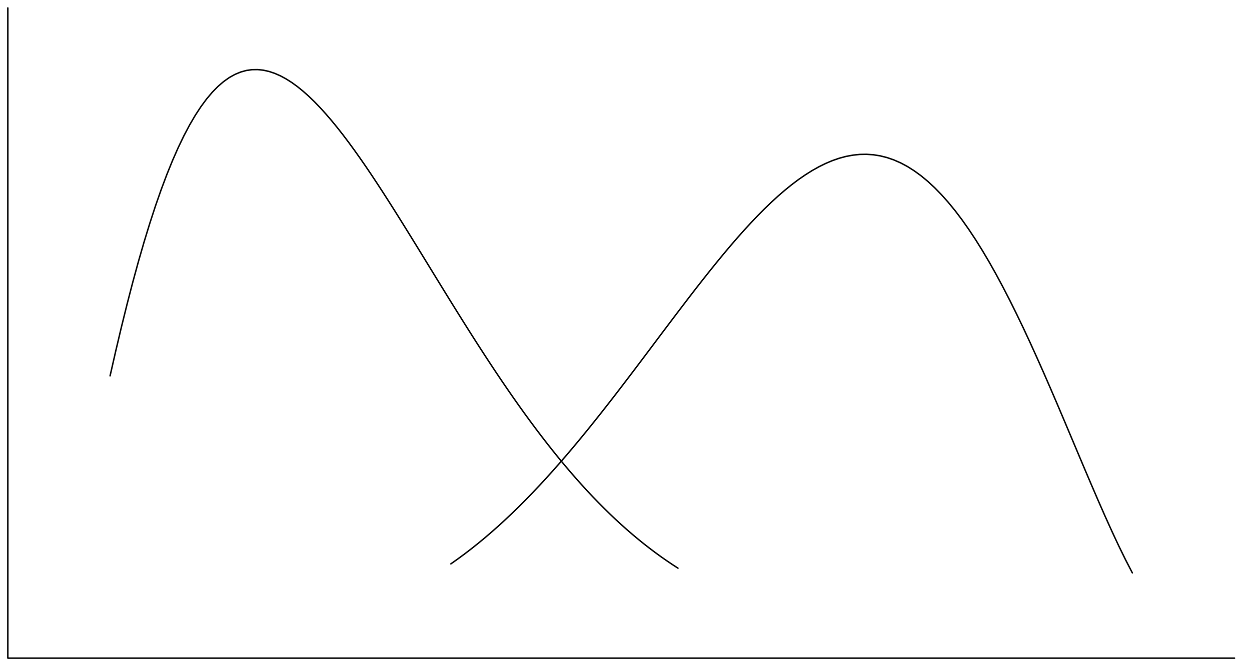
Discussion

- **Overlap**



Discussion

- **Disjoint support**



Discussion

- **Choice of conditioning variables**
  - For the CIA to be credible:
    - Observable variables of past (especially income)
    - On current variables or variables measured after introduction of treatment: be careful not to use variables that may be affected by the programme
    - Difference estimation: if we have data on past before treatment, we use difference before and after (like DD)
  - IMPORTANT: For the common support assumption to hold, explanatory variables must not over-explain being treated because that might make matching with beneficiaries that are too close impossible
  - Introducing too many variables may result in biased estimates

Discussion

- **Recap for matching**
  - Estimating procedure
    - Selection of control group
    - Choice of conditioning variables (observable variables)
      - It is not about describing the probability of treatment as accurately as possible but ‘simply’ determining the variables necessary to obtain the independence property
    - Estimating the propensity score (Probit, logit)
    - Determining the common support
      - E.g. eliminating observations whose propensity score is close to 1 or 0
    - Checking balance score: have variables *X* been balanced in both groups?
    - Matching on basis of score
    - Calculation of precision
      - May be costly in computing time

Discussion

- **Comments**
  - Implementing these methods may be complex
    - Setting many technical parameters: bandwidth, degrees, number of twins
    - No clear rules
  - Possible extensions to multi-treatment (but slightly more complex interpretation)

QUIZ 16

- Situation: You have to evaluate the impact of a support scheme for high-school pupils in danger of dropping out. The scheme is open to pupils, who decide whether or not to take part. You calculate the propensity score for each pupil, that is, the probability of taking part depending on several observable characteristics: their academic results, age, sex and parents' occupation.
- Case 1: the propensity score is 40–55% in the treated group and 35–50% in the untreated group.
- Case 2: the propensity score is 45–55% in the treated group and 35–50% in the untreated group.
- Given this information, what can you expect about case 2 compared to case 1?
    - Outcomes will have less scope in case 2 than case 1.
    - Both cases will yield the same outcomes because there is common support in both cases.
    - The analysis will suffer from greater bias in case 2 than case 1.

QUIZ 17

- Situation: You have to evaluate the impact of a support scheme for high-school pupils in danger of dropping out. The scheme is open to pupils, who decide whether or not to take part. You calculate the propensity score for each pupil, that is, the probability of taking part depending on several observable characteristics: their academic results, age, sex and parents' occupation.
- We assume that participation in the scheme depends largely on pupils' motivation, which also impacts their success when leaving school.
- In this case, does the matching method reduce bias compared to a linear regression of success when leaving school on participation in the scheme, by controlling for different possible combinations of academic results, age, sex and parents' occupation?
  - Yes
  - No

Conclusion

- Methods of selection on observables are often the only feasible solution for conducting an evaluation
- They include a large choice of techniques: check the sensitivity of results to different choices of specification
- They do not solve the problems of unobservables