האם בתי ספר פרטיים טובים יותר מבתי ספר ציבוריים?
הערכת המצב באירלנד בשטות מחקר תצפיתני

דני פפרמן, האוניברסיטה העברית ואוניברסיטת סאות'המפטון, אנגליה,
הלשכה המרכזית לסטטיסטיקה, ישראל

ויקטוריה לנדסמן, אוניברסיטת טורונטו, קנדה

הכינוס ה- 11, האגודה הישראלית לפסיכומטריקה, פברואר 2015

מאמר:

Are private Schools better than public schools? Appraisal for Ireland by methods for observational studies.
***The Annals of Applied Statistics*, 2011, 5, 1726–1751.**

The problem of observational Studies

Assignment of units to treatments (and possibly the sample selection) not under control.

Estimation of treatment effects based on the observations under the various treatments without adjustment can be biased, since the units exposed to the various treatments could differ in important unknown characteristics related to the response.

- **How can we estimate the true treatment effects?**

Notation and definitions

Population: U=1…*N*. Every unit $i \in$ U potentially exposed to each of *m* treatments with outcomes $y_i^t, t = 1...m$.

→ "Counterfactual approach"

Target parameters:

$$\mu^{p,t} = \sum_{i=1}^{N} y_i^t / N, \quad \mu^t = \sum_{i=1}^{N} E(y_i^t \mid x_i) / N \quad \ldots \boldsymbol{t}\textbf{=1…}\boldsymbol{m},$$

Or contrasts between the means, e.g., ($\mu^1 - \mu^2$)

⇓

Average Treatment Effect (ATE)

Notation and definitions (cont.)

Sample S of *n* observational units obtained with (generally unknown) probabilities $\pi_i = \Pr(i \in S)$.

In practice, every unit *j* ∈ S exposed to one treatment with probability, $p_j^t = \Pr[T(j) = t \mid j \in S]$ **;** $\sum_{t=1}^{m} p_j^t = 1$

$$\Downarrow$$

$$P(i \in S, T(i) = t) = \pi_i \times p_i^t = q_i^t$$

After assignment, $S = S^1 \cup, ..., \cup S^T$ **;** $S^t = \{i \mid i \in S, T(i) = t\}$.

Problem revisited

For an uncontrolled observational study,

$$f_{S^t}(y_i^t \mid x_i) = f(y_i^t \mid x_i, \boldsymbol{i \in S^t}) \neq f_U(y_i^t \mid x_i) = f(y_i^t \mid x_i, \boldsymbol{i \in U}).$$

$f(y_i^t \mid x_i, i \in U)$**= *distribution* of** y_i^t **if every unit** $i \in$**U is exposed to treatment *t* (‘population model’).**

In particular,

$$E_{S^t}(y_i^t \mid i \in S^t) \neq E_U(y_i^t \mid i \in U)$$

unless under “strong ignorability”.

Some methods in common use

- **Assume the availability of auxiliary (*X*) or instrumental (*Z*) variables that control the assignment bias;** $Y_{S^t} \perp T \mid X \text{ or } Z$ **.**

Suppose m=2, 1= treatment, 0=control.

A- Regression methods:

Assumption: $E_{S^t}(y_i^t \mid x_i, i \in S^t) = E_U(y_i^t \mid x_i) = r^t(x_i)$**, *t*=0,1**

$$\Downarrow$$

$$\hat{ATE} = \frac{1}{n}\sum_{i=1}^{n}[\hat{r}^1(\bar{x}) - \hat{r}^0(\bar{x})].$$

Some methods in common use (cont.)

B- Matching methods

Denote $J_M^t(i) =$ **set of *M* closest matches in S^t for unit *i* $\in$S^{1-t} based on *x*.**

Observe: $\hat{y}_i^{1-t} = y_i^{1-t}$ **; impute:** $\hat{y}_i^{1-t} = \frac{1}{M} \sum_{j \in J_M^t(i)} y_j^t$

$$\hat{ATE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i^1 - \hat{y}_i^0)$$

- **Regression methods impute** $\hat{y}_i^t = \hat{r}^t(x_i)$ **for all** i**.**
- **Matching often based on propensity scores (next slide).**

Some methods in common use (cont.)

C- Methods based on propensity scores (PS)

Assumption: $P(T_i = 1 \mid y_i, x_i) = P(T_i = 1 \mid x_i) = e(x_i)$ **(PS)**

Estimated based on sampled units, (*logistic*, *probit*…)

C1- $$\hat{ATE} = \sum_{i=1}^{n} \frac{T_i y_i}{\hat{e}(x_i)} \Big/ \sum_{i=1}^{n} \frac{T_i}{\hat{e}(x_i)} - \sum_{i=1}^{n} \frac{(1-T_i) y_i}{[1-\hat{e}(x_i)]} \Big/ \sum_{i=1}^{n} \frac{(1-T_i)}{[1-\hat{e}(x_i)]}$$ **, (Hajek)**

T_i=1 $\rightarrow$treatment, T_i=0 $\rightarrow$control

D- Methods based on Econometric models (Heckman, Instrumental variables,…), Double robustness…

Limitations of existing methods

Problem: $f_{S^t}(y_j^t \mid x_j) = f(y_j^t \mid x_j, j \in S^t) \neq f_p(y_j^t \mid x_j)$.

Existing methods assume the availability of auxiliary (instrumental) variables that control the assignment bias.

$$Y_{S^t} \perp T \mid X \text{ or } Z$$

- **Availability of *X* or *Z* not guaranteed.**
- **Difficult (impossible?) to test the assumptions.**

Different approach

Problem: $f_{S^t}(y_j^t \mid x_j) = f(y_j^t \mid x_j, j \in S^t) \neq f_p(y_j^t \mid x_j)$.

So, why not use the sample distribution for inference?

$$f_{S^t}(y_j^t \mid x_j) = \frac{\Pr(j \in S^t \mid y_j^t, x_j) f_p(y_j^t \mid x_j)}{\Pr(j \in S^t \mid x_j)}$$ **(Bayes Theorem)**

$f_p(y_i^t \mid x_i)$ - ***pdf*** **under 'ignorable' assignment.**

- $f_{S^t}(y_j^t \mid x_j) = f_p(y_j^t \mid x_j) \Leftrightarrow \Pr(j \in S^t \mid y_j^t, x_j) = \Pr(j \in S^t \mid x_j)$

↓

propensity scores

Inference based on the sample distribution

$$f_{S^t}(y_j^t \mid x_j) = \frac{\Pr(j \in S^t \mid y_j^t, x_j)\, f_p(y_j^t \mid x_j)}{\Pr(j \in S^t \mid x_j)}$$

Model $\Pr(j \in S^t \mid y_j^t, x_j; \alpha_t)$ **and** $f_p(y_j^t \mid x_j; \theta_t)$**. Estimate unknown parameters by maximum likelihood or other methods.**

Estimation of treatment effects

A- $\hat{\mu}^t = N^{-1} \sum_{j=1}^{N} \hat{E}_p(y_j^t \mid x_j)$

B- $\hat{\mu}^{pt} = \sum_{j \in S^t} (y_j^t / \hat{p}_j^t) / \sum_{j \in S^t} (1/\hat{p}_j^t)$ **;** $\hat{p}_j^t = \hat{\Pr}(j \in S^t \mid y_j^t, x_j)$**.**

- **Use of B does not require knowledge of *x* for units** $i \notin S$**.**

Two fundamental questions

1. Is the sample distribution identifiable?

2. Can we test the model?

- **General conditions guaranteeing model identifiability established in the paper and in articles by other researchers.**

Can we test the model?

Yes, why not? The sample data are independent observations from the *sample distribution* , $f_{S^t}(y \mid x)$**.**

Example: Kolmogorov-Smirnov test statistic.

$$KS = \sup_{y^t} \left| F^t_{EMP}(y^t) - \hat{F}_{S^t}(y^t; \hat{\theta}_t, \hat{\alpha}_t) \right|; \quad \hat{F}_{S^t}(y^t; \hat{\theta}_t, \hat{\alpha}_t) = \frac{1}{n_t} \sum_{j \in S^t} \hat{F}^t_j(y^t \mid x_j)$$

$$F^t_{EMP}(\mathrm{y}^t) = \frac{1}{n_t} \sum_{j \in S^t} \mathrm{I}(y^t_j \le y^t)$$ **=empirical distribution;** $n_t = \#(S^t)$**.**

$\hat{\theta}_t, \hat{\alpha}_t$ **= Estimates of sample distribution parameters.**

- **Critical values of the KS statistic can be obtained by parametric Bootstrap (*Babu and Rao*, 2004).**

So, does the method work ???

Programme for International Student Assessment (PISA).

- **Collects information on children's proficiency in maths, science and reading + family and school characteristics.**
- **32 countries; school children aged 15.**
- **Survey waves every 3 years (first wave in year 2000).**
- **In the present application we compare children's scores in Math between private and public schools in Ireland.**

Sample sizes: private=1256 ; public=702.

- **Data analyzed by Vandenberghe and Robin (2004).**

Covariates

Constant, gender, father education, socio-economic index, index of home educational resources, socio-economic index at school level.

- **Outcome values and continuous covariates standardized**

Instrumental variable (for existing methods)

School location: 1 if school in big city

0 otherwise

- **Used in other studies.**

Model fitted

Population model (normal)

$y_i^t \sim \mathrm{N}\left(x_i'\beta_t, \sigma_t^2\right)$, ***t*=0,1;** $x_i' \rightarrow$ **covariates.**

Assignment probabilities (logistic)

$$P(i \in S^t \mid x_i, y_i^t) = \frac{\exp(c_t + \boldsymbol{\delta_t} y_i^t + x_i'\gamma_t)}{1+\exp(c_t + \boldsymbol{\delta_t} y_i^t + x_i'\gamma_t)},$$ ***t*=0,1 (NOT PS)**

➢ **Instrumental variable Z_i included among the covariates.**

Estimation of model parameters in private schools

Assignment (logistic)

Coefficient	const	δ_1	Gen.	F.edu	S.E.I	H.E.R	S.E.S	S.loc
Estimate	-2.95	0.49	0.77	0.04	-0.12	3.16	0.09	1.13
Std error	1.30	0.21	0.13	0.12	0.07	0.20	0.07	0.13

Population (normal)

Parameter	σ_1	const	Gen.	F.edu	S.E.I	H.E.R	S.E.S	S.loc
Estimate	0.83	6.09	-0.20	0.18	0.16	0.39	0.21	- 0.09
Std error	0.02	0.07	0.05	0.05	0.03	0.09	0.02	0.06

➢ Supports the use of propensity scores and Inst. variables.

Estimation of model parameters in public schools

Assignment (logistic)

Coefficient	const	δ_0	Gen.	F.edu	S.E.I	H.E.R	S.E.S	S.loc
Estimate	13.88	-2.02	-0.76	0.17	0.40	-2.57	0.27	-1.63
Std error	2.90	0.39	0.18	0.23	0.12	0.30	0.11	0.24

Population (normal)

Parameter	σ_0	const	Gen.	F.edu	S.E.I	H.E.R	S.E.S	S.loc
Estimate	1.10	6.89	0.17	0.10	0.16	1.35	0.30	0.23
Std error	0.07	0.14	0.08	0.09	0.04	0.20	0.04	0.15

- Use of propensity scores questionable.

Estimation of population means by type of school

Private School			Public School	
	$\hat{\mu}^1 = \overline{x}'\hat{\beta}^1$	$\hat{\mu}^1_{DR}$	$\hat{\mu}^0 = \overline{x}'\hat{\beta}^0$	$\hat{\mu}^0_{DR}$
Estimate	**6.10**	**6.09**	**7.05**	**6.91**
Std	**0.05**	**0.06**	**0.15**	**0.12**

Estimation of ATE for Ireland

New method

Method	$\hat{\Delta} = \hat{\mu}^1 - \hat{\mu}^0$	$\hat{\Delta}_{DR}$
Estimator	**-0.95**	**- 0.82**
Std error	**0.16**	**0.13**

Model diagnostics

Statistic	K-S	U^t	Moran	Max_{ps}
Private schools	**0.0218** **PV=0.13**	**-0.48** **PV=0.64**	**- 0.464** **PV=0.64**	**0.04812**
Public schools	**0.040** **PV=0.17**	**-1.24** **PV=0.22**	**- 0.165** **PV=0.87**	

$$\boldsymbol{Max_{PS}} = \max_{j \in S} \left| 1 - \sum_{t=1}^{m} \hat{\Pr}(j \in S^t \mid x_j) \right|$$

$$\Pr(j \in S^t \mid x_j) = \int \Pr(j \in S^t \mid y_j^t, x_j)\, f_p(y_j^t \mid x_j)\, dy_j^t \quad \textbf{(PS)}.$$

Under correct model $\rightarrow \sum_{t=1}^{m} \Pr(j \in S^t \mid x_j) = 1$ **for every** ***j*****.**

Estimation of ATE by new and existing methods

New method

Method	$\hat{\Delta} = \hat{\mu}^1 - \hat{\mu}^0$	$\hat{\Delta}_{DR}$
Estimator	-1.05	- 0.89
Std error	0.25	0.23

Existing methods

Method	$\bar{y}^1 - \bar{y}^0$	Reg.	PS Match	PS Hajek	Double Robust	Instrument
Estimate	0.36	0.12	0.21	0.16	0.17	- 0.61
Std error	0.05	0.05	0.05	0.05	0.05	0.24

Simulations

- Generated 400 data sets from the models fitted to Ireland and verified that model parameters and ATE are estimated unbiasedly and that the test statistics perform properly.

- Generated 400 data sets from a model with error terms generated from a t_4-distribution, but fitted the model that assumes normal error terms. The test statistics indicated wrong model specification.