25th EMS

Oslo, Wednesday 27.7.2005

The effect of university studies on job opportunities:

an application of the principal strata approach to causal inference

Leonardo Grilli
grilli@ds.unifi.it

Fabrizia Mealli
mealli@ds.unifi.it

Department of Statistics

University of Florence

Outline

- Scope and motivation
- The principal strata framework
- Model specification
- ML results
- Future work

Scope and motivation

AIM: assessing the relative effectiveness of two degree programmes with respect to employment

- 1992 cohort of freshmen of the University of Florence
- Two degree programmes: Economics and Political Science
- Employment: binary indicator for having a permanent job about two years after degree

Scope and motivation

Naif approach: compare the employment rates for the graduates

But this is not fair, because the two degree programmes might "select" the individuals in a different way (e.g. one d.p. might be more easy in general or for students with certain features)

(issue is relevant: in our data the graduation rate after 8 years is around 25%)

If the graduates of the two d.p. differ for some unobserved features which are related with the occupational chances then a comparison based only on graduates yields biased results ⇒ need to take into account the graduation process

Methodological approach

- Assessing relative effectiveness is an example of causal inference
- Causal inference can be approached in many ways: here we follow the **potential outcomes** approach (Rubin)
- Within such an approach we exploit the idea of **principal stratification**, since in our application there is a relevant intermediate variable (graduation) between the treatment variable (chosen degree prog.) and the outcome variable (employment)

Frankgakis C.E. & Rubin D.B. (2002) Principal stratification in causal inference, *Biometrics*, 58: 21-29.

Barnard J., Frangakis C.E., Hill J.L. & Rubin D.B. (2003) Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City, *JASA*, 98: 299-323.

Main variables

Treatment variable Z:

$$Z = \begin{cases} 1 & \text{if enrolled in Economics} \\ 0 & \text{if enrolled in Political Science} \end{cases}$$

- Z is called "treatment" just to conform to the literature on causal inference
- No active vs. placebo $\rightarrow$ values of Z on an equal footing
- No randomisation $\rightarrow$ possible confounders (so covariates are important for unconfoundedness)

Main variables

Intermediate variable S:

$$S = S(z) = \begin{cases} 1 & \text{if graduated when } z \\ 0 & \text{if not graduated when } z \end{cases}$$

S is the observed version of the potential variables S(0), S(1)

Outcome variable Y:

$$Y = Y(z) = \begin{cases} 1 & \text{if employed (after graduation) when } z \\ 0 & \text{if not employed (after graduation) when } z \end{cases}$$

Y is the observed version of the potential outcomes Y(0), Y(1)

For our purposes Y is defined only when S=1

The principal strata framework

In our case both Z and S are binary $\rightarrow$ 4 possible strata

Z	L=GG	L=GN	L=NG	L=NN
1	G	G	N	N
0	G	N	G	N

G=Graduated

N=Not graduated

Principal strata are defined by values of the two potential versions of the intermediate var. S (counterfactual): e.g. **GN** are the students who become **G**raduate if enrolled in Economics and **N**ot graduate if enrolled in Political Sc.

By definition principal strata are not influenced by Z (nor S)

The membership indicator of the principal stata L is a categorical latent (i.e. unobserved) covariate $\Rightarrow$ need for ***latent class models***

The principal strata framework

Relationship between observed and latent groups

Observed group $O(Z, S^{obs})$	Z_i	S_i^{obs}	Y_i^{obs}	Latent group L_i (principal stratum)
$O(1,1)$	1	1	in {0,1}	GG or GN
$O(1,0)$	1	0	not defined	NG or NN
$O(0,1)$	0	1	in {0,1}	GG or NG
$O(0,0)$	0	0	not defined	GN or NN

Every observed group is a mixture of principal strata

Relevant parameters

Probabilities of the principal strata: π_{GG}, π_{GN}, π_{NG}, π_{NN}

e.g. probability to be a student who become **G**raduate if enrolled in Economics and **N**ot graduate if enrolled in Political Science

Probabilities of employment: $\gamma_{1,GG}$, $\gamma_{0,GG}$, $\gamma_{1,GN}$, $\gamma_{0,NG}$

e.g. probability to be employed for a student who (i) become **G**raduate if enrolled in Economics and **N**ot graduate if enrolled in Political Science and (ii) actually enrolled in Economics

Causal effect of degree prog. on employment in the GG group: $\gamma_{1,GG} - \gamma_{0,GG}$

Type of analysis

Principal stratification is the conceptual framework for the application of various statistical methods:

- Non parametric methods ($\Rightarrow$ bounds)
- Model-based methods ($\Rightarrow$ point estimates)
  - ML
  - Bayesian

Our paper (Grilli & Mealli 2005, submitted and available on request) shows how to obtain large-sample non parametric **bounds** using minimal assumptions and ML **point estimates** adding further assumptions

In this talk I show the ML results

Likelihood

$$L\left(\boldsymbol{\theta} \mid \mathbf{Z}, \mathbf{S}^{obs}, \mathbf{Y}^{obs}, \mathbf{X}\right) =$$

$$\prod_{i \in O(1,1)} \left\{ \pi_{GG:i} \left[\left(\gamma_{1,GG:i}\right)^{Y_i^{obs}} \left(1-\gamma_{1,GG:i}\right)^{1-Y_i^{obs}} \right] + \pi_{GN:i} \left[\left(\gamma_{1,GN:i}\right)^{Y_i^{obs}} \left(1-\gamma_{1,GN:i}\right)^{1-Y_i^{obs}} \right] \right\}$$

$$\times \prod_{i \in O(1,0)} \left\{ \pi_{NG:i} + \pi_{NN:i} \right\}$$

$$\times \prod_{i \in O(0,1)} \left\{ \pi_{GG:i} \left[\left(\gamma_{0,GG:i}\right)^{Y_i^{obs}} \left(1-\gamma_{0,GG:i}\right)^{1-Y_i^{obs}} \right] + \pi_{NG:i} \left[\left(\gamma_{0,NG:i}\right)^{Y_i^{obs}} \left(1-\gamma_{0,NG:i}\right)^{1-Y_i^{obs}} \right] \right\}$$

$$\times \prod_{i \in O(0,0)} \left\{ \pi_{GN:i} + \pi_{NN:i} \right\}$$

Various models can be built by specifying submodels for the π's and the γ's

Model specification

Probabilities of the principal strata: $\pi_{GG}, \pi_{GN}, \pi_{NG}, \pi_{NN}$

Principal strata submodel: multinomial logit

Probabilities of employment: $\gamma_{1,GG}, \gamma_{0,GG}, \gamma_{1,GN}, \gamma_{0,NG}$

Outcome submodel: 4 separate logit models

Principal strata are latent classes

$\Rightarrow$ the model is a special instance of a *latent class model*

(with restrictions, since a given individual can belong to only two of the four classes)

Data

A. **Administrative database** of the 1992 cohort of freshmen enrolled in *Economics* (1068 students) and *Political Science* (873 students)

B1-B3. Three **census surveys** on the occupational status of the graduates of the University of Florence of years 1998 to 2000

datasets A and B1-B3 are merged

Available covariates: Female, Residence in Florence, Gymnasium, High grade, Late enrollment

covariates are important since the treatment is not randomized!

ML inference

- ✓ Maximization algorithm: quasi-Newton with a BFGS update of the Cholesky factor of the approximate Hessian
- ✓ Software: SAS proc NLMIXED

- Principal strata submodel ⇒ 18 parameters
- Outcome submodel ⇒ 9 parameters

Overall 27 parameters

Some parameters of the Principal strata submodel (a multinomial logit) have highly negative estimates and huge standard errors

⇒ for certain values of the covariates some principal strata are empty so some constraints are needed (the final model has 8 constraints)

Principal strata submodel results

- the size of GG stratum varies a lot with the covariates, from a minumum of 1.1% (students with weak background) to a maximum of 62.2%
- for most covariate patterns the GN and NG strata (i.e. students able to graduate in only one d. p.) are very small (but for students with weak background they are larger then the GG stratum)
- the higher graduation rate of Economics is originated by the students with a weak background $\Rightarrow$ orientation policies should be designed especially for this kind of students

Outcome submodel results

- the level of the probability of being employed varies a lot with the covariates
- in the GG stratum the causal effect on employment (modelled as costant across the covariate patterns) is about 15% (significant at 5%)
- students with a weak background have little chances of being GG, so for them the above causal effect has little relevance

Final remarks

- Future work
  - Alternative model specifications
  - Sensitivity analysis
  - More than two degree programmes
- Questions and further material
  - Email: **grilli@ds.unifi.it**
  - Web: **www.ds.unifi.it/grilli**

Thanks for your attention!

Additional slides

Principal strata submodel

$$\pi_{GG:i} = \frac{\exp(\eta^{\pi}_{GG:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

$$\pi_{GN:i} = \frac{\exp(\eta^{\pi}_{GN:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

$$\pi_{NG:i} = \frac{\exp(\eta^{\pi}_{NG:i})}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

$$\pi_{NN:i} = \frac{1}{1+\exp(\eta^{\pi}_{GG:i})+\exp(\eta^{\pi}_{GN:i})+\exp(\eta^{\pi}_{NG:i})}$$

Multinomial logit specification

$$\eta^{\pi}_{GG:i} = \alpha^{\pi}_{GG} + \boldsymbol{\beta}^{\pi}_{GG}{}'\mathbf{x}_i$$

$$\eta^{\pi}_{GN:i} = \alpha^{\pi}_{GN} + \boldsymbol{\beta}^{\pi}_{GN}{}'\mathbf{x}_i$$

$$\eta^{\pi}_{NG:i} = \alpha^{\pi}_{NG} + \boldsymbol{\beta}^{\pi}_{NG}{}'\mathbf{x}_i$$

With 5 covariates there are $3+3\times5=18$ parameters

Principal strata submodel results

	Initial model		*Final model*	
Number of parameters	27		21	
Deviance (-2$logL$)	2231.8		2231.8	
Principal strata submodel (π's)				
α^{π}_{GG}	-4.403	(0.449)	-4.402	(0.448)
α^{π}_{GN}	-2.644	(0.749)	-2.647	(0.752)
α^{π}_{NG}	-3.206	(0.836)	-3.207	(0.835)
$\beta^{\pi}_{GG,gymnasium}$	1.275	(0.157)	1.275	(0.157)
$\beta^{\pi}_{GN,gymnasium}$	-5.757	(n.a.)	$-\infty$	
$\beta^{\pi}_{NG,gymnasium}$	-15.041	(n.a.)	$-\infty$	
$\beta^{\pi}_{GG,high\_grade}$	1.204	(0.146)	1.205	(0.146)
$\beta^{\pi}_{GN,high\_grade}$	1.113	(0.653)	1.113	(0.652)
$\beta^{\pi}_{NG,high\_grade}$	-8.092	(114.022)	$-\infty$	
$\beta^{\pi}_{GG,regular\_enrolment}$	2.024	(0.425)	2.023	(0.425)
$\beta^{\pi}_{GN,regular\_enrolment}$	-0.012	(0.788)	-0.009	(0.792)
$\beta^{\pi}_{NG,regular\_enrolment}$	-8.140	(64.473)	$-\infty$	
$\beta^{\pi}_{GG,female}$	0.117	(0.137)	0.117	(0.137)
$\beta^{\pi}_{GN,female}$	-0.617	(0.753)	-0.622	(0.755)
$\beta^{\pi}_{NG,female}$	0.988	(1.112)	0.991	(1.111)
$\beta^{\pi}_{GG,Florence}$	0.280	(0.144)	0.280	(0.144)
$\beta^{\pi}_{GN,Florence}$	-13.499	(559.599)	$-\infty$	
$\beta^{\pi}_{NG,Florence}$	-10.353	(533.855)	$-\infty$	

Outcome submodel

$$\gamma_{1,GG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{1,GG:i})}$$

$$\gamma_{0,GG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{0,GG:i})}$$

$$\gamma_{1,GN:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{1,GN:i})}$$

$$\gamma_{0,NG:i} = \frac{1}{1+\exp(-\eta^{\gamma}_{0,NG:i})}$$

Separate logit specifications

$$\eta^{\gamma}_{1,GG:i} = \alpha^{\gamma}_{1,GG} + \boldsymbol{\beta}^{\gamma}\,'\mathbf{x}_i$$

$$\eta^{\gamma}_{0,GG:i} = \alpha^{\gamma}_{0,GG} + \boldsymbol{\beta}^{\gamma}\,'\mathbf{x}_i$$

$$\eta^{\gamma}_{1,GN:i} = \alpha^{\gamma}_{1,GN} + \boldsymbol{\beta}^{\gamma}\,'\mathbf{x}_i$$

$$\eta^{\gamma}_{0,NG:i} = \alpha^{\gamma}_{0,NG} + \boldsymbol{\beta}^{\gamma}\,'\mathbf{x}_i\,.$$

With 5 covariates there are 4+5=9 parameters

Outcome submodel results

	Initial model		*Final model*	
Number of parameters		27		21
Deviance (-2*logL*)		2231.8		2231.8
Outcome submodel (γ's)				
$\alpha^{\gamma}_{1,GG}$	1.257	(1.240)	1.262	(1.241)
$\alpha^{\gamma}_{0,GG}$	-1.357	(1.561)	-1.365	(1.568)
$\alpha^{\gamma}_{1,GN}$	0.593	(1.185)	0.596	(1.185)
$\alpha^{\gamma}_{0,NG}$	0.498	(1.057)	0.484	(1.058)
$\beta^{\gamma}_{gymnasium}$	-0.405	(0.374)	-0.410	(0.374)
$\beta^{\gamma}_{high\_grade}$	-0.035	(0.262)	-0.036	(0.263)
$\beta^{\gamma}_{regular\_enrolment}$	-0.933	(0.979)	-0.932	(0.979)
β^{γ}_{female}	0.072	(0.272)	0.070	(0.272)
$\beta^{\gamma}_{Florence}$	0.106	(0.333)	0.104	(0.333)
Causal effect $\alpha^{\gamma}_{1,GG} - \alpha^{\gamma}_{0,GG}$	0.664	(0.301)	0.666	(0.301)

Estimated probabilities (per cent) for some covariate patterns

Probability	**00000**	**00100**	**00110**	**00101**	**01100**	**10100**	**11100**	**11111**
$\pi_{GG:i}$	1.1	8.0	9.1	10.9	20.3	24.9	52.5	62.2
$\pi_{GN:i}$	6.3	6.0	3.3	0.0	14.0	0.0	0.0	0.0
$\pi_{NG:i}$	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\pi_{NN:i}$	89.0	86.0	87.6	89.1	65.7	75.1	47.5	37.8
$\gamma_{1,GG:i}$	77.9	58.2	59.9	60.7	57.3	48.0	47.1	51.5
$\gamma_{0,GG:i}$	64.5	41.7	43.4	44.2	40.8	32.2	31.4	35.3
$\gamma_{1,GN:i}$	61.9	39.0	40.7	41.5	38.1	29.8	29.0	32.8
$\gamma_{0,NG:i}$	20.3	9.1	9.7	10.0	8.9	6.3	6.1	7.1
Causal effect $\gamma_{1,GG:i} - \gamma_{0,GG:i}$	13.5	16.5	16.5	16.4	16.5	15.8	15.7	16.2

Note: the pattern $(x_1, x_2, x_3, x_4, x_5)$ stands for $Gymnasium = x_1, High\ grade = x_2, Regular\ enrolment = x_3,$ $Female = x_4, Florence = x_5$.