Workshop on causality

Ripa (Perugia), 29-30 April 2005

ESTIMATING DIRECT AND INDIRECT EFFECTS: DISCUSSION OF SOME APPROACHES

Leonardo Grilli
grilli@ds.unifi.it

Fabrizia Mealli
mealli@ds.unifi.it

Università degli Studi di Firenze
Dipartimento di Statistica "Giuseppe Parenti"

Outline

- Direct/indirect effects: 3 approaches
- Causal effects with intermediate variables
- Principal strata and mapping variables
- Surrogates and truncation "by death"
- An application: effectiveness of degree programmes
- Causal effects via SEM

Direct and indirect causal effects

Z S Y

Z affects Y both directly and indirectly through an intermediate variable S, and there is an interest or a necessity to disentangle the two

Main problem: S is a post-treatment variable, so conditioning on it may lead to bias

Direct and indirect causal effects

Z S Y

Two special cases:

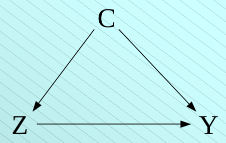
- Surrogate (informal definition):
  - S is a surrogate for the effect of Z on Y when all the effect of Z on Y is mediated by S, i.e. there is no direct effect (moreover, for practical use S should be a good predictor of Y)
- Truncation "by death"
  - When Y is not defined for certain values of S (in such a case the missingness of Y is due to non-existence rather than to non-response) ⇒ problematic since the effect of the treatment is undefined for certain values of a post-treatment variable

Three approaches to causal inference: solutions for direct/indirect effects

Potential outcomes	Causal DAGs	Structural equations
Rubin	Lauritzen, Pearl, Spirtes, Dawid	Heckman
↓	↓	↓
Principal strata	**Natural direct eff.** **Mapping variables**	Systems of eqs.

Causal DAGs: Identification of causal effects

Causal DAG: conditional distributions are stable under interventions

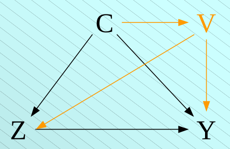


Z = treatment
Y = response
C = covariates

Criterion for identification: **BACK-DOOR**

Causal DAGs: Identification of causal effects

BACK-DOOR and CONFOUNDING:



Z = treatment

Y = response

C = covariates

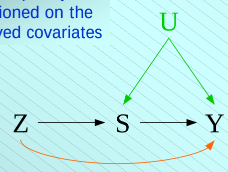
V = unobs. confounder

If V affects both Z and Y back-door criterion is violated!

Need to remove the arrow V → Z (e.g. randomisation, sufficient covariates) or V → Y (e.g. sufficient covariates)

Intermediate variables

Graph implicitly conditioned on the observed covariates



Z = treatment

Y = response

S = intermediate

U = unobs. variables

arrow Z → Y: direct effect

Assumption: conditionally on the observed covariates there are *no unobserved confounders*, i.e. no arrow U → Z

Causal effects with an intermediate variable

BACK-DOOR with intermediate S:

If the conditions of the back-door criterion are satisfied (as in the previous DAG) the *total* effect of Z on Y is identified

However, if for some values of the intermediate S the response Y is not defined (*truncation* "*by death*") the *total* effect of Z on Y **is not defined for every individual**

Causal effects with an intermediate variable

The arrow Z → Y represents the direct effect of Z on Y, i.e. the effect non mediated by S

Truncation "by death": when the arrow Z → Y is present, it may be that even the direct effect is defined *only* for a subset of individuals, e.g. those individuals having S=1 irrespective of the value of Z

Surrogates: only individuals having S(0)=S(1) can provide evidence of a direct effect of Z on Y

Causal effects with an intermediate variable

Under truncation by death the direct effect of Z on Y is defined *only* for a subset of individuals, e.g. having S=1 irrespective of the value of Z

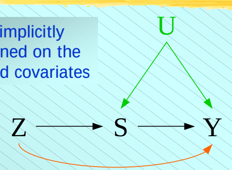
Such subset of individuals is not observable

how can a DAG represent truncation by death?

S. Lauritzen suggests to use **mapping variables**

Principal strata and mapping variables

Graphs implicitly conditioned on the observed covariates



Z = treatment

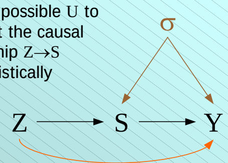
Y = response

S = intermediate

U = unobs. var.

σ = mapping var. $\chi_Z \to \chi_S$

Coarsest possible U to represent the causal relationship Z→S deterministically



Given σ, knowledge of Z implies knowledge of S

It is not specified how S and Y depend on U!

Causal inference with principal strata

Principal causal effect of Z on Y:

p(Y(1)) vs. p(Y(0)) for the individuals of a given principal strata

or, in terms of the graph,

$p(Y \mid Z=1,\sigma)$ vs. $p(Y \mid Z=0,\sigma)$ for a given σ

Note the use of $p(Y \mid Z,\sigma)$ instead of $p(Y \mid Z, S, U)$

Average causal effects across principal strata are nonsense

Under truncation by death the causal effect is defined only for some principal strata (e.g. the stratum with S(0)=S(1)=1)

Surrogates and direct effects

S. Lauritzen uses mapping variables to represent in a DAG the concepts of direct principal effect and principal surrogate

Since σ is a partition of U it follows that (provided Z is binary)

Strong surrogate	$Y \perp Z \mid S, U$
⇓	
Principal surrogate	$Y \perp Z \mid S, \sigma$

Direct principal effect	$Y \not\perp Z \mid S, \sigma$
⇓	
Direct effect	$Y \not\perp Z \mid S, U$

An application: effectiveness of degree programmes

Grilli & Mealli (2005)

An application: effectiveness of degree programmes

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

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

An application: effectiveness of degree programmes

Why is it not fair to compare employment for the graduated students only?

- Because it is possible that the two degree programmes "select" the individuals in a different way (e.g. one d.p. is more easy in general or for students with certain features)

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 gradutated students yields biased results

An application: effectiveness of degree programmes

Treatment variable Z:

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

- No active vs. placebo → values of Z on an equal footing
- No randomisation → possible confounders (so covariates are important for ignorability)

An application:
effectiveness of degree programmes

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)

Response variable Y:

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

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

For our purposes Y is defined only when S=1

An application:
effectiveness of degree programmes

In our case both Z and S are dichotomous → 4 possible strata

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

G=Graduated

N=Not graduated

Principal strata are defined by the values of the two potential versions of the intermediate variable 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.

Principal strata are not influenced by Z (nor S)

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

An application:
effectiveness of degree programmes

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

An application:
effectiveness of degree programmes

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

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

An application:
effectiveness of degree programmes

Some approaches:

- Non parametric analysis (⇒ bounds)
- Model-based analysis
  - Likelihood
  - Bayesian

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

An application:
effectiveness of degree programmes

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

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

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

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

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

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

An application: effectiveness of degree programmes

Model-based inference: our specification

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

Multinomial logit model

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

4 separate logit models

Since the principal strata are latent classes, the model is in fact a latent class model (though with some restrictions, since the observed values of the treatment and intermediate variable are incompatible with certain latent classes)

Direct and indirect effects via Structural Equations Models

Model specification

Z ⟶ S ⟶ Y

Binary variable generated by dichotomisation of a latent continuous response:

$$S_i = I\left(\tilde{S}_i > 0\right) \qquad Y_i = I\left(\tilde{Y}_i > 0\right)$$

Latent responses generated by the following SEM with two linear equations:

$$\tilde{S}_i = \alpha^{(S)} + \beta^{(S)} Z_i + \varepsilon_i^{(S)}$$
$$\tilde{Y}_i = \alpha^{(Y)} + \gamma^{(Y)} S_i + \beta^{(Y)} Z_i + \varepsilon_i^{(Y)}$$

$$E(\varepsilon_i^{(S)}) = E(\varepsilon_i^{(Y)}) = 0 \qquad corr(\varepsilon_i^{(S)}, \varepsilon_i^{(Y)}) = \rho \neq 0$$

Variance of error terms not identifiable, but fixed (e.g. 1 for probit, $\pi^2/3$ for logit)

Special cases

$$\tilde{S}_i = \alpha^{(S)} + \beta^{(S)} Z_i + \varepsilon_i^{(S)}$$
$$\tilde{Y}_i = \alpha^{(Y)} + \gamma^{(Y)} S_i + \beta^{(Y)} Z_i + \varepsilon_i^{(Y)}$$

S is a *dummy endogenous variable* ⇒ its effect on *Y* cannot be unbiasedly estimated using only the equation for *Y*

$\beta^{(Y)} = 0$ ⇒ *Y* is independent of *Z* given *S*

S is called a *surrogate* for the effect of *Z* for *Y*. It may seems that in such a case there is no direct effect of *Z* on *Y*, but this conclusion is questionnable when unobserved variables are present (Frangakis & Rubin, 2002)

$\gamma^{(Y)} = 0$ ⇒ *S* canceals from the equation on Y

the effect of *Z* on *Y* is entirely direct and (given unconfoundedness of *Z*) such an effect can be unbiasedly estimated using only the equation on Y

Sample selection vs. Truncation by death

- Y is observed or defined only when S assumes certain values, e.g. when S=1

$$\tilde{S}_i = \alpha^{(S)} + \beta^{(S)} Z_i + \varepsilon_i^{(S)}$$
$$\tilde{Y}_i = \alpha^{(Y)} + \beta^{(Y)} Z_i + \varepsilon_i^{(Y)}$$

  - Sample selection ⇒ Y unobserved when S=0 though it is defined (so observable in principle)
  - Truncation by death ⇒ Y unobserved when S=0 because it is not defined
- under truncation by death the coefficient of Z on Y *is defined only for* (and thus estimable only from) the unobservable subset of individuals for whom *S*=1 irrespective of Z (so the extension to all individuals is meaningless)

Truncation by death

- As in sample selection, due to non random missingness of *Y*, unbiased estimation requires to fit both equations simultaneously
- As in sample selection, non parametric identification requires an instrumental variable (i.e. a variable affecting *S* but not *Y*) but this does not solve the problems of definition
- The issue is clear using principal strata, since the potential outcomes Y(0),Y(1) (and so the corresponding causal effect) are defined only in the principal strata with S(0)=S(1)=1

Principal strata

$$\{S_i = 1\} \Leftrightarrow \left\{\varepsilon_i^{(S)} > -\left(\alpha^{(S)} + \beta^{(S)} Z_i\right)\right\}$$

$\beta^{(S)} \geq 0$ no loss of generality since the coding of the binary variables is arbitrary

$$S_i = \begin{cases} 1 & \text{if} \quad \varepsilon_i^{(S)} > -\alpha^{(S)} & \text{P.Stratum } S(0) = S(1) = 1 \\ Z_i & \text{if} \quad -\left(\alpha^{(S)} + \beta^{(S)}\right) < \varepsilon_i^{(S)} \leq -\alpha^{(S)} & \text{P.Stratum } S(0) = 0, S(1) = 1 \\ 0 & \text{if} \quad \varepsilon_i^{(S)} \leq -\left(\alpha^{(S)} + \beta^{(S)}\right) & \text{P.Stratum } S(0) = S(1) = 0 \end{cases}$$

With this specification there is monotonicity: S=1-Z is impossible

Further stratification

$$\{Y_i = 1\} \Leftrightarrow \left\{\varepsilon_i^{(Y)} > -\left(\alpha^{(Y)} + \gamma^{(Y)} S_i + \beta^{(Y)} Z_i\right)\right\}$$

$\beta^{(Y)} \geq 0$ no loss of generality since the coding of the binary variables is arbitrary

$\gamma^{(Y)}$ Various scenarios to be considered

Let us consider: 1. Standard case 2. truncation by death

Standard case

Values of Y(0),Y(1) for latent classes defined by model errors $\left(\gamma^{(Y)} \geq \beta^{(Y)} \geq 0\right)$

(0,0)	(0,1)	(1,1)	(1,1)	(1,1)
(0,0)	(0,1)	(0,1)	(0,1)	(1,1)
(0,0)	(0,0)	(0,0)	(0,1)	(1,1)

Row boundaries ($\varepsilon_i^{(S)}$): $-\alpha^{(S)}$, $-\left(\alpha^{(S)} + \beta^{(S)}\right)$

Column boundaries ($\varepsilon_i^{(Y)}$): $-(\alpha^{(Y)} + \gamma^{(Y)} + \beta^{(Y)})$, $-(\alpha^{(Y)} + \gamma^{(Y)})$, $-(\alpha^{(Y)} + \beta^{(Y)})$, $-\alpha^{(Y)}$

Each row is a principal stratum

Each cell is a further partitioning such that also Y is deterministic, so in each cell the causal effect of Z on Y is constant

With $\gamma^{(Y)} \geq \beta^{(Y)}$ no cell can have a negative effect

Standard case

In general the total **Average Causal Effect** of Z on Y is

$$\begin{aligned} & E(Y \mid Z=1) - E(Y \mid Z=0) \\ & = \sum_c \left[E(Y \mid Z=1, C=c) - E(Y \mid Z=0, C=c)\right] P(C=c) \\ & = \sum_{c \in \mathcal{C}^+} P(C=c) - \sum_{c \in \mathcal{C}^-} P(C=c) \end{aligned}$$

Standard case

When $\gamma^{(Y)} \geq 0$ the total ACE of Z on Y is the sum of

$P\left(\varepsilon_i^{(S)} > -\alpha^{(S)}, -(\alpha^{(Y)} + \gamma^{(Y)} + \beta^{(Y)}) < \varepsilon_i^{(Y)} \leq -(\alpha^{(Y)} + \gamma^{(Y)})\right)$ direct eff.

$P\left(-(\alpha^{(S)} + \beta^{(S)}) < \varepsilon_i^{(S)} \leq -\alpha^{(S)}, -(\alpha^{(Y)} + \gamma^{(Y)} + \beta^{(Y)}) < \varepsilon_i^{(Y)} \leq -\alpha^{(Y)}\right)$ Mixed direct /indirect eff.

$P\left(\varepsilon_i^{(S)} \leq -(\alpha^{(S)} + \beta^{(S)}), -(\alpha^{(Y)} + \beta^{(Y)}) < \varepsilon_i^{(Y)} \leq -\alpha^{(Y)}\right)$ direct eff.

The mixed direct /indirect eff. can be decomposed under assumptions

When $\beta^{(Y)} = 0$ the direct effects vanish

Standard case

Values of Y(0),Y(1) for latent classes defined by model errors $\beta^{(Y)} \geq 0 \geq \gamma^{(Y)}$, $\beta^{(Y)} + \gamma^{(Y)} \leq 0$

(0,0)	(0,0)	(0,1)	(0,1)	(1,1)
(0,0)	(0,0)	(1,0)	(1,1)	(1,1)
(0,0)	(0,1)	(1,1)	(1,1)	(1,1)

Row boundaries ($\varepsilon_i^{(S)}$): $-\alpha^{(S)}$, $-\left(\alpha^{(S)} + \beta^{(S)}\right)$

Column boundaries ($\varepsilon_i^{(Y)}$): $-(\alpha^{(Y)} + \beta^{(Y)})$, $-\alpha^{(Y)}$, $-(\alpha^{(Y)} + \gamma^{(Y)} + \beta^{(Y)})$, $-(\alpha^{(Y)} + \gamma^{(Y)})$

Here one cell has a negative effect, so total ACE may be negative

Truncation by death

Values of Y(0),Y(1) for latent classes defined by model errors

(0,0)	(0,1)	(1,1)
(*,0)	(*,1)	(*,1)
(*,*)	(*,*)	(*,*)

$-\alpha^{(S)}$ $-\left(\alpha^{(S)}+\beta^{(S)}\right)$ $\varepsilon_i^{(S)}$

$\varepsilon_i^{(Y)}$ $-(\alpha^{(Y)}+\beta^{(Y)})$ $-\alpha^{(Y)}$

Here the causal effect is defined only in the principal stratum of the individuals having *S*=1 irrespective of *Z*. Therefore ACE is

$$P\left(-(\alpha^{(Y)}+\beta^{(Y)}) < \varepsilon_i^{(Y)} \le -(\alpha^{(Y)}) \;\; | \;\; \varepsilon_i^{(S)} > -\alpha^{(S)} \right)$$

Final remarks

- Even if a lot of similarities and correspondences can be found, different approaches lead to different identifying assumptions (behavioral, distributional, functional)
- Some (distributional) assumptions obscure others that relate to possibly latent but meaningful subgroups of individuals
- Often the traslation of a scientific theory into a statistical model is more direct using principal stratification rather than structural equations
- Principal stratification prevents extrapolation from one latent group to another: if this is an important issue or not, strongly depends on the empirical context

Effectiveness of degree programmes: some details of the application

Grilli & Mealli (2005)

Data

A. *Administrative database* of the 1992's cohort of freshmen enrolled in the degree programmes in Economics (Economia e Commercio) and Political Science (Scienze Politiche) of the University of Florence

B1-B3. Three *census surveys* on the occupational status of the graduates of the University of Florence of years 1998, 1999 and 2000, respectively

Datasets A and B1-B3 are merged

Data

1941 freshmen belong to the examined 1992's cohort: 1068 in *Economics* and 873 in *Political Sciences*. By the end of the year 2000 the status of the students is the following:

Degree Programme	Dropped	Graduated	Still enrolled	Total
Economics	545 51.03%	270 25.28%	253 23.69%	1068
Political Sciences	532 60.94%	176 20.16%	165 18.90%	873

Data

After the merge with the survey data the situation is:

Degree Programme	Graduated	Interviewed	Permanent job
Economics	270	186 68.89%*	96 51.61%**
Political Sciences	176	99 56.25%*	36 36.36%**

* Interviewed/Graduated **Permanent job/Interviewed

All interviewed graduates responded to the question on job status. Apart from 21 students who graduated before 1998 (out of the target of the surveys), almost all missing interviews are due to **missing contact**

Data

Covariate	*Economics (n=1068)*	*Political Science (n=873)*
Female	0.41	0.54
Residence in Florence	0.23	0.31
Gymnasium	0.34	0.45
Late enrollment	0.06	0.22
High grade	0.37	0.25

Covariates are important since the treatment is not randomized!

Model

Principal strata submodel (π's)

Multinomial logit specification

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

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

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

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

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

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

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

With 5 covariates there are 3+3×5=18 parameters

Model

Outcome submodel (γ's)

Separate logit specifications

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

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

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

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

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

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

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

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

With 5 covariates there are 4+5=9 parameters

Maximum likelihood inference

- Model has 18+9=27 parameters
- The treatment and the 5 covariates lead to 128 theoretical sample proportions
- The available sample proportions are 99

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

Maximum likelihood inference

- Some parameters of the Principal strata submodel (π's) have
  - ✓highly negative estimates and
  - ✓huge standard errors

⬇

for certain values of the covariates some principal strata are empty

⬇

some constraints are needed
(the final model has 8 constraints)

Principal strata submodel results

- The estimated *proportion of students belonging to the GG group* varies a lot with the covariates, from a minumum of 1.1% (*students with weak background*) to a maximum of 62.2%
- the *proportions of students belonging to the GN and NG groups* (i.e. the students able to graduate in only one degree programme) are very small (but for some covariate patterns the GN and NG groups are larger then the GG group)

Principal strata submodel results

- the two degree programmes have a *differential causal effect on the probability of graduation only for students having a weak background*. Orientation policies should then be designed especially for this kind of students.

Outcome submodel results

- the *causal effect in the GG group* (on the logit scale) is estimated as 0.666 (s.e. 0.301, significant at 5%) corresponding to a difference of about 15% in the probabilities of employment
- the reliability and also the substantive importance of the causal effect depends on the *size of the GG stratum*: for example, the causal effect in the *GG* group for *students having a weak background* has little relevance

Outcome submodel results

- The level of the probability of being employed varies a lot with the covariates:
  - ✓ 47.1% to 77.9% for Economics
  - ✓ 31.4% to 64.5% for Political Science

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 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 covariates' 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 $Gymnasium = x_1, High\ grade = x_2, Regular\ enrolment = x_3, Female = x_4, Florence = x_5$.