4th German Stata Users' Group Meeting
University of Mannheim, Germany
31 March 2006

MANCHESTER 1824

Intervention evaluation using gllamm

Andrew Pickles

(University of Manchester)

GLLAMM and gllamm

- GLLAMM is a modelling framework most fully elaborated in the book

  Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Chapman & Hall/CRC Press. Boca Raton, FL.

- gllamm is a software implementation that is capable of fitting very many of the models with the GLLAMM framework.

  - Rabe-Hesketh, S., Pickles, A. and Taylor, C. (2000). sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin* **53**, 47-57.

  - Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* **2**, 1-21.

- gllamm now consists of a model fitting program, and post-estimation and simulation programs gllapred and gllasim.

- gllamm and gllamm manual, datasets and other information are available from *www.gllamm.org*

GLLAMM and gllamm

What do GLLAMM and gllamm let you do?

GLLAMM helps you to understand and gllamm allows you to analyse the effects of covariates and the structure of covariance (multivariate normal and discrete mixture) among sets of measures that may be of different kinds (continuous, count, nominal, ordered, ranked, censored)

GLLAMM and gllamm

This includes for any response type:

- variance components (including frailty models)
- random coefficient and growth curve models
- factor analysis
- structural equation models
- latent class models
- selection models
- non-ignorable non-response
- multilevel versions of the above

GLLAMM and gllamm

This generality is gained at some expense.

Speed: for any 'standard' analysis a specialist program will run more quickly.

Speed is improving as the result of the efforts of StataCorp, the gllamm team (Sophia Rabe-Hesketh, Andrew Pickles and Anders Skrondal) and as computers improve.

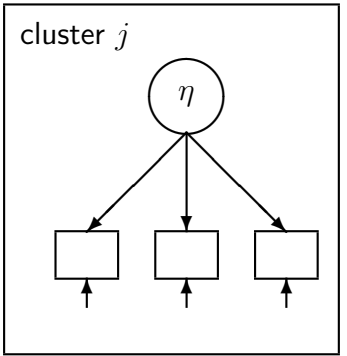
Model set-up: some more complex models can require careful prior data manipulation. The writing of wrapper programs that do this for you for particular model types is in progress.

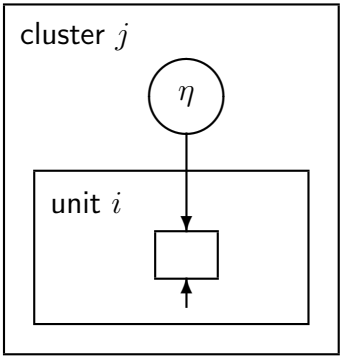
Generalized linear mixed models

We can add random effects into any GLM

- Clustered or 'two-level' data: level-1 units i nested in level-2 clusters j
  - Repeated measurements on patients
  - Twins in families
- Unobserved between-cluster covariates (or unobserved heterogeneity)
  $\Longrightarrow$ Dependence between units ij and $i'j$ in the same cluster j
- Include a cluster-specific random intercept η_j in the linear predictor

$$\nu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \eta_j$$





Note:

☞ frames indicate 'level'

☞ ◯ encloses latent variables

☞ □ surrounds observed var.

☞ $\longrightarrow$ represents a regression

Random coefficient models in GLLAMM

- One covariate multiplies each latent variable,

$$\eta_m^{(l)} z_{m1}^{(l)} \quad (\lambda_{m1}^{(l)} = 1)$$

- e.g. Latent growth curve model for individuals j (level 2) observed at times t_{ij}, $i = 1, \cdots, n_j$ (level 1)

$$\text{Linear predictor: } \nu_{ij} = \beta_1 + \beta_2 t_{ij} + \eta_{1j}^{(2)} + \eta_{2j}^{(2)} t_{ij}$$

β_1, β_2 : mean intercept and slope

$\eta_{1j}^{(2)}$, $\eta_{2j}^{(2)}$: random deviations of unit-specific intercepts and slopes from their means

Generalized random coeff. model in GLLAMM

$$\nu = \mathbf{x}'\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_m^{(l)\prime} \boldsymbol{\lambda}_m^{(l)}$$

For identification, $\lambda_{m1}^{(l)} = 1$

- Fixed part: $\mathbf{x}'\boldsymbol{\beta}$ as usual
- Random part:
  - $\eta_m^{(l)}$ is mth latent variable at level l, $m = 1, \cdots, M_l$, $l = 2, \cdots, L$
    Can be a factor or a random coefficient
  - $\mathbf{z}_m^{(l)}$ are variables and $\boldsymbol{\lambda}_m^{(l)}$ are parameters
  - Unless regressions for the latent variables are specified, latent variables at different levels are independent whereas latent variables at the same level may be dependent

gllamm syntax for estimating GLMMs

gllamm [*varlist*] [if *exp*] [in *range*] , i(*varlist*) [nrf(*numlist*)
eqs(*eqnames*) offset(*varname*) family(*family*) link(*link*) eform
nip(*numlist*) adapt from(*matrix*) ...]

i(*varlist*) $L-1$ variables identifying the hierarchical, nested clusters, from level 2 to L, e.g., `i(pupil class school)`.

nrf(*numlist*) $L-1$ numbers specifying the numbers of latent variables M_l at each level.

eqs(*eqnames*) $M = \sum M_l$ equations for the $\mathbf{z}_m^{(l)\prime}\boldsymbol{\lambda}_m^{(l)}$ multiplying each latent variable. Constants must be explicitly included in the equation definition.

family(*family*), link(*link*) and eform as for glm.

offset(*varname*) variable in fixed part with regression coefficient set to 1.

nip(*numlist*) numbers of quadrature points for each latent variable (total M), a single number meaning that all values are the same.

adapt adaptive quadrature will be used.

from(*matrix*) passes starting values to gllamm – use skip if matrix contains extra parameters and copy if column and equation names not right.

Syntax examples: linear predictor

- Two-level growth curve model (occasions in subjects)

$$\text{Linear predictor: } \nu_{ij} = \beta_1 + \beta_2 t_{ij} + \eta_{1j}^{(2)} + \eta_{2j}^{(2)} t_{ij}$$

```
gen cons=1
eq int: cons
eq slope: time
gllamm y time, i(subject) nrf(2) eqs(int slope) ...
```

- Three-level growth curve model (occasions in subjects in centres)

$$\text{Linear predictor: } \nu_{ijk} = \beta_1 + \beta_2 t_{ijk} + \eta_{1jk}^{(2)} + \eta_{2jk}^{(2)} t_{ijk} + \eta_{1k}^{(3)} + \eta_{2k}^{(3)} t_{ijk}$$

```
gllamm y time, i(subject centre) nrf(2 2) /*
   */ eqs(int slope int slope) ...
```

gllapred syntax for prediction

gllapred *varname* [if *exp*] [in *range*] [, xb u linpred mu
marginal us(*varname*) outcome(#) above(#) ···]

xb fixed part of linear predictor returned in *varname*.

u posterior means and standard deviations of latent variables returned in *varname*m1, *varname*s1, *varname*m2, etc.

ustd same as u but divided by approximate sampling standard deviation.

linpred linear predictor (with posterior means of latent variables) returned in *varname*.

mu mean response $\mathrm{E}[g^{-1}(\nu)]$ returned in *varname*. By default expectation w.r.t. posterior distribution.

marginal marginal or population average mean (expectation w.r.t. prior distribution).

us(*varname*) expectation conditional on latent variables being equal to the values in *varname*1, *varname*2, etc.

outcome(#) with mlogit link, probability that the response equals #.

above(#) with ordinal links, probability that response exceeds #.

gllasim syntax for simulation

gllasim *varname* [if *exp*] [in *range*] [, u us(*varname*) from(*matrix*) ⋯]

By default, responses are simulated for the model just estimated and returned in *varname*.

u latent variables are simulated and returned in *varname*p1, *varname*p2, etc.

us(*varname*) response variables are simulated for latent variables equal to *varname*1, *varname*2, etc.

from(*matrix*) causes responses/latent variables to be simulated from the model just estimated in gllamm but with parameter values in matrix.

Growth and trajectory models: treatment of depression

Postnatal depression example

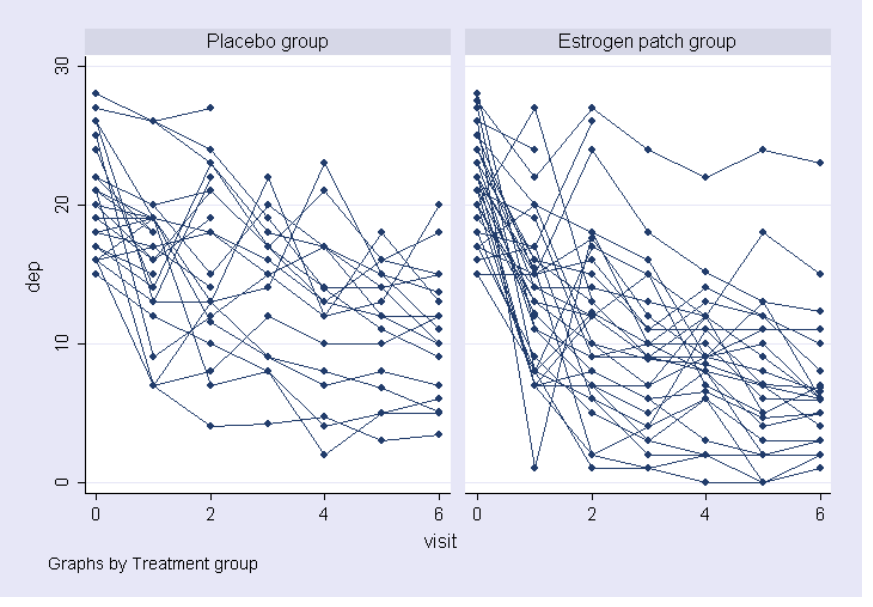
The data look like

```
use depress7.dta, clear
list, clean
```

	subj	visit	group	dep
1.	1	0	Placebo group	18
2.	1	1	Placebo group	17
3.	1	2	Placebo group	18
4.	1	3	Placebo group	15
5.	1	4	Placebo group	17
6.	1	5	Placebo group	14
7.	1	6	Placebo group	15
8.	2	0	Placebo group	27
9.	2	1	Placebo group	26
10.	2	2	Placebo group	23
...				
349.	59	0	Estrogen patch group	17
350.	59	1	Estrogen patch group	15
351.	60	0	Estrogen patch group	22
352.	60	1	Estrogen patch group	7
353.	60	2	Estrogen patch group	12
354.	60	3	Estrogen patch group	15
355.	61	0	Estrogen patch group	26
356.	61	1	Estrogen patch group	24

Postnatal depression example

```
sort group subj visit
twoway (connected dep visit, connect(ascending)), by(group)
```



Depression example: growth curve model

Response at time t of individual i, y_{it}, is given by:

$$y_{it} = \underbrace{\alpha + \beta t}_{\text{fixed part}} + \underbrace{\eta_{it}}_{\substack{\text{random} \\ \text{effects}}} + \underbrace{e_{it}}_{\substack{\text{occasion} \\ \text{specific error}}}$$

where

$$\eta_{it} = u_{1i} + u_{2i}t$$

and $(u_{1i}, u_{2i}) \sim$ bivariate normal.

In the standard growth curve model the random effects for slope and intercept are allowed to be correlated.

Postnatal depression example

Bivariate random effects model

```
gen con=1
eq int: con
eq slope: visit
xi: gllamm dep i.group*visit, i(subj) nrf(2) eqs(int slope) adapt
```

```
...
number of level 1 units = 356
number of level 2 units = 61

Condition Number = 28.96942

gllamm model

log likelihood = -1041.133
```

dep	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Igroup_1	-1.653089	1.035749	-1.60	0.110	-3.683121	.3769425
visit	-1.526425	.2091052	-7.30	0.000	-1.936264	-1.116587
_IgroXvisi~1	-.5464383	.2660811	-2.05	0.040	-1.067948	-.0249289
_cons	19.2888	.7769387	24.83	0.000	17.76603	20.81157

```
Variance at level 1
------------------------------------------------------------------------------
14.4725 (1.2985379)

Variances and covariances of random effects
------------------------------------------------------------------------------
***level 2 (subj)

    var(1): 8.9642528 (2.9576111)
    cov(2,1): .38745363 (.54299217) cor(2,1): .25252183

    var(2): .26261984 (.16961806)
------------------------------------------------------------------------------
```

Postnatal depression example

Compare random intercept model with random coefficient model by using Likelihood Ratio Test

Model 1: random intercept model

```
xi: gllamm dep i.group*visit, i(subj) adapt
    ... log likelihood = -1045.7117

estimates store model1          /* store estimates in model1 */
```

Model 2: Random coefficient model

```
xi: gllamm dep i.group*visit, i(subj) nrf(2) eqs(int slope) adapt
    ... log likelihood = -1041.133
```

Likelihood ratio test:

```
lrtest model1 .          /* compare model1 with current */
    (log-likelihoods of null models cannot be compared)

    likelihood-ratio test                          LR chi2(2)  =      9.16
    (Assumption: model1 nested in .)               Prob > chi2 =    0.0103
```

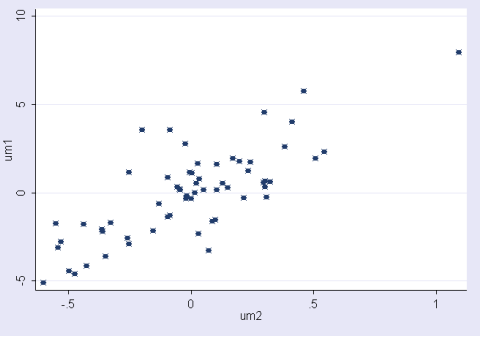
Note:

- ☞ Likelihood ratio test not valid since null hypothesis on boundary of parameter space
- ☞ Snijders and Bosker (1999) and others suggest dividing p-value by 2

Postnatal depression example

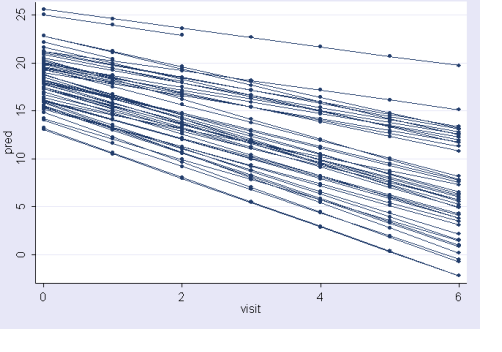
- Obtaining estimates of the random effects for individual deviations for intercepts and slopes

```
gllapred u, u
twoway (scatter um1 um2)
```



- Obtaining estimates of individual predicted values (trajectories)

```
gllapred pred, mu
sort subj visit
twoway (connected pred visit, msymbol(smcircle) /*
    */ connect(ascending))
```



bmatrix option in gllamm

bmatrix(*matrix*) specifies a matrix B of regression coefficients for the dependence of the latent variables on other latent variables. The matrix must be upper diagonal and have number of rows and columns equal to the total number of random effects.

Depression example by using bmatrix

An alternative setup is to let one of the random effects be regressed upon the other:

$$\begin{aligned} \eta_1 &= 0\eta_1 + \beta\eta_2 + \zeta_1 \\ \eta_2 &= 0\eta_1 + 0\eta_2 + \zeta_2 \end{aligned}$$

where ζ_1 and ζ_2 are uncorrelated.

```
constraint 1 [sub1_2_1]_cons=0

matrix b=(0,1 \ 0,0)

xi: gllamm dep i.group*visit, i(subj) nrf(2) nip(8) eqs(int slope) /*
       */ bmatrix(b) nocorrel adapt
```

Depression example by using bmatrix

Output

```
...
log likelihood = -1041.133021837493
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Igroup_1	-1.653089	1.035749	-1.60	0.110	-3.68312	.3769416
visit	-1.526425	.2091052	-7.30	0.000	-1.936264	-1.116587
_IgroXvisi~1	-.5464382	.2660812	-2.05	0.040	-1.067948	-.0249287
_cons	19.2888	.7769384	24.83	0.000	17.76603	20.81157

```
 Variance at level 1
------------------------------------------------------------------------------
  14.472499 (1.2985371)

Variances and covariances of random effects
------------------------------------------------------------------------------
***level 2 (subj)
    var(1): 8.392612 (4.101821)
    cov(2,1): 0 (0) cor(2,1): 0
    var(2): .26262034 (.16961689)

 B-matrix:
------------------------------------------------------------------------------
    B(1,2): 1.4753391 (2.6476786)
------------------------------------------------------------------------------
```

☞ This gives the same likelihood, fixed effects estimates. The variance of the slope is 0.2626 as before, but the variance of the intercept is now given by $Var(\zeta_1) + b^2 Var(\zeta_2) = 8.3926 + 1.4753^2 * 0.2626 = 8.964$ (the same value as before).

Latent trajectory models

Response at time t of individual i, y_{it}, is given by a **growth model**:

$$y_{it} = \underbrace{\alpha + \beta t}_{\text{fixed part}} + \underbrace{\eta_{it}}_{\substack{\text{random} \\ \text{effects}}} + \underbrace{e_{it}}_{\substack{\text{occasion} \\ \text{specific error}}}$$

The η_{it}'s are represented by discrete trajectory classes c with probability π_c:

$$(\eta_{it} \mid c) = e_{1c} + e_{2c}t,$$

where

- e_{1c} is the trajectory origin or intercept for class c
- e_{2c} is the trajectory slope for class c
- Prevalence of trajectory class c is π_c
- $\sum_{k=1}^{C} \pi_k e_{1k} = 0$ and $\sum_{k=1}^{C} \pi_k e_{2k} = 0$

Latent trajectory models

We will hereafter consider three models:

Model 1: *unconditional trajectory classes and unconditional class probabilities*

Model 2: *unconditional trajectory classes and conditional class probabilities*

☞ *We allow probability* π_{ic} *that subject* i *belongs to latent class* c *to depend on covariates* $\boldsymbol{x_i}$ *through a multinomial logit model. For example, if we consider just one covariate* x_i*:*

$$\pi_{ic} = \frac{\exp(\gamma_{0c} + \gamma_{1c} x_i)}{\sum_{k=1}^{C} \exp(\gamma_{0k} + \gamma_{1k} x_i)},$$

where the γ_{0k}*'s and the* γ_{1k}*'s are parameters.*

Model 3: *conditional trajectory classes and unconditional class probabilities:*

$$y_{it} = \alpha + \beta x_i + \beta x_i t + \eta_{it} + e_{it}$$

☞ *Covariate effects included in fixed part of the model*

☞ *Classes now represent groups having accounted for covariate differences*

Postnatal depression example

- **Latent trajectory model (1)**: unconditional trajectory classes and unconditional class probabilities

```
gen cons=1
eq int: cons
eq slope: visit
gllamm dep visit, i(subj) nrf(2) eq(int slope) ip(f) trace nip(2)
```

```
...
```

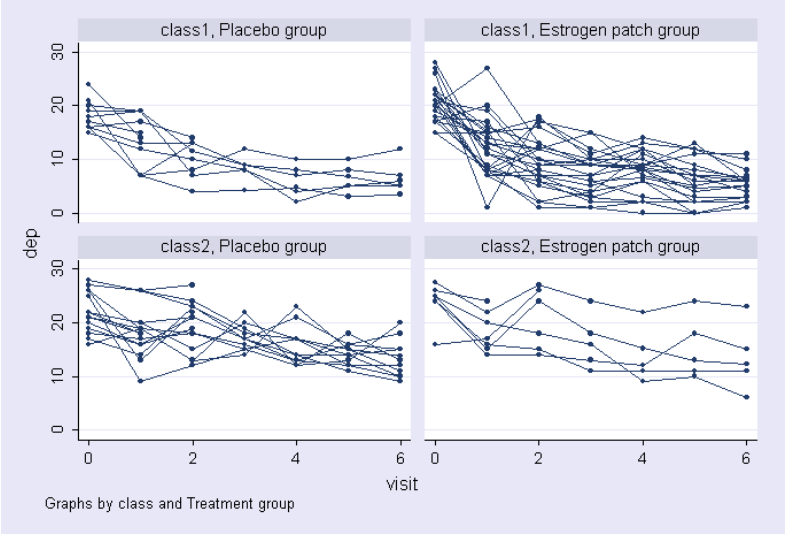
dep	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
visit	-1.898491	.1363647	-13.92	0.000	-2.165761	-1.631221
_cons	18.38703	.4981955	36.91	0.000	17.41058	19.36347

```
Variance at level 1
------------------------------------------------------------------------------
  19.139691 (1.4643147)

Probabilities and locations of random effects
------------------------------------------------------------------------------
***level 2 (subj)

    loc1: -1.9586, 2.933
  var(1): 5.7444392

    loc2: -.31928, .47814
cov(2,1): .9364582
  var(2): .15266137
    prob: 0.5996, 0.4004

    log odds parameters
    class 1
    _cons: .40381744 (.31445191)
------------------------------------------------------------------------------
```

Postnatal depression example

Now assign women to classes and look at what distinguishes one class from another.

```
preserve
gllapred prob, p
gen class=cond(prob1>prob2,1,2)
label define classl 1 "class1"  2 "class2"
label values class classl
sort class subj visit
twoway (connected dep visit, msymbol(smcircle) connect(ascending)), by(class group)
```



Postnatal depression example

Test for association of class assignment with treatment:

```
tab class group if visit == 0, chi2
```

class	Treatment group: Placebo g	Estrogen	Total
class1	11	27	38
class2	16	7	23
Total	27	34	61

```
Pearson chi2(1) =   9.5815   Pr = 0.002
```

```
restore
```

☞ **Note**: we reject the null hypothesis that `class` and `group` are independent.

Postnatal depression example

Let's model treatment differences in latent class probabilities directly.

➠ **Latent trajectory model (2)**: unconditional trajectory classes and conditional class probabilities

```
eq clprob: group
gllamm dep visit, i(subj) nrf(2) eq(int slope) peqs(clprob) ip(f) trace nip(2)
...
```

dep	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
visit	-1.639986	.176207	-9.31	0.000	-1.985345	-1.294626
_cons	19.66	.6530511	30.10	0.000	18.38004	20.93996

```
Variance at level 1
------------------------------------------------------------------------------
19.192753 (1.4748225)

Probabilities and locations of random effects
------------------------------------------------------------------------------
***level 2 (subj)

    loc1: -3.1888, 1.6681
  var(1): 5.3192671

    loc2: -.54866, .28701
cov(2,1): .91522481
  var(2): .15747215
    prob: 0.3435, 0.6565

    log odds parameters
    class 1
    group: 2.1258399 (.70207624)
    _cons: -.64795694 (.46781989)
```

☞ treatment effect on class assignment

Postnatal depression example

- **Latent trajectory model (3)**: conditional trajectory classes and unconditional class probabilities

```
gen gpvisit=group*visit
gllamm dep visit gpvisit, i(subj) nrf(2) eq(int slope) ip(f) trace nip(2)
...
```

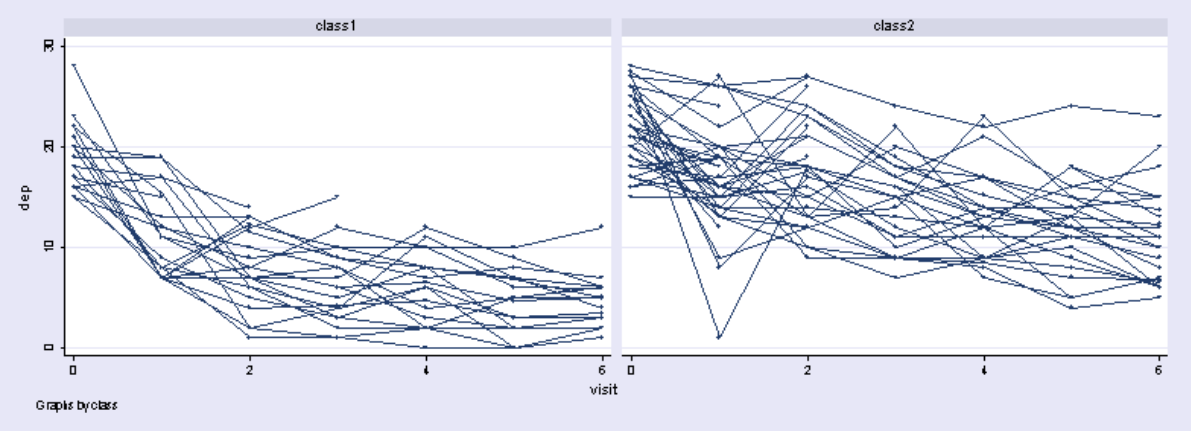
dep	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
visit	-1.424514	.1655199	-8.61	0.000	-1.748927	-1.100101
gpvisit	-.7501039	.1692819	-4.43	0.000	-1.08189	-.4183175
_cons	18.36341	.4986261	36.83	0.000	17.38612	19.3407

```
Variance at level 1
------------------------------------------------------------------------------
  18.927176 (1.4531254)

Probabilities and locations of random effects
------------------------------------------------------------------------------
***level 2 (subj)

    loc1: -3.0379, 1.9312
  var(1): 5.8667044

    loc2: -.31252, .19867
cov(2,1): .60354323
  var(2): .06209013
    prob: 0.3886, 0.6114

    log odds parameters
    class 1
    _cons: -.45301726 (.32825506)
------------------------------------------------------------------------------
```

Postnatal depression example

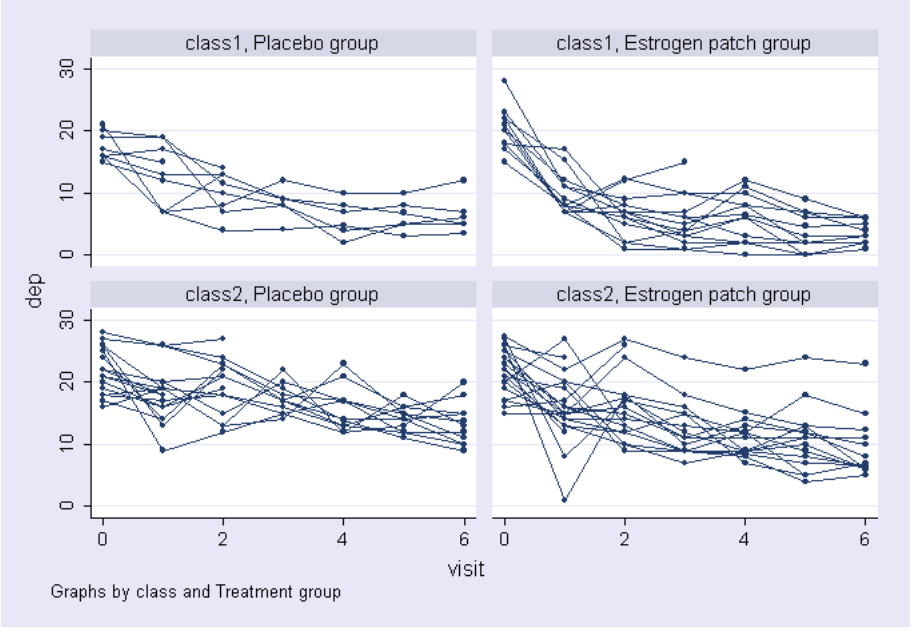
Posterior probabilities:

```
gllapred prob, p
gen class=cond(prob1>prob2,1,2)
label define classl 1 "class1"  2 "class2"
label values class classl
sort class subj visit
twoway (connected dep visit, msymbol(smcircle) connect(ascending)), /*
    */ by(class) ysize(8) xsize(20)
```



Postnatal depression example

```
twoway (connected dep visit, msymbol(smcircle) connect(ascending)), /*
     */ by(class group)
```



Postnatal depression example

Test for association of class assignment with treatment:

```
tab class group if visit == 0, chi2
```

class	Treatment group: Placebo g	Estrogen	Total
class1	9	14	23
class2	18	20	38
Total	27	34	61

```
Pearson chi2(1) =   0.3941   Pr = 0.530
```

☞ **Note:** As expected, we accept the null hypothesis of independence since the treatment effect has already been accounted for in the fixed part and the latent classes relate to variation around the fixed part.

Instrumental variables and CACE estimation

Trials that go wrong

- In many trials treatment assignment does not fully determine treatment exposure. Non-compliance results in other factors also influencing exposure.

- It cannot be assumed that those other factors are not selective. In other words some aspects of exposure may be associated with confounders.

- Nonetheless can exploit random assignment as an instrumental variable, to identify part of the variation in exposure that is uncorrelated with confounders.

The ODIN trial

- Psychological treatment for depression in primary care.
- Eight centres throughout Europe.
- Participants (N=427) allocated to receive psychological treatment or treatment as usual.
- Only about half of those patients allocated to treatment actually take up the offer.
- Loss to follow-up is associated with **non-compliance**.

The ODIN trial

Compliance rates vary across Centres – from 40 to 74%

	Treatment Group		Control
	C	NC	
1 Eire	6 (40%)	9	23
2 Spain	12 (63%)	7	11
3 Finland	17 (74%)	6	24
4 Finland	20 (71%)	8	22
5 Norway	22 (52%)	20	25
6 Norway	17 (47%)	19	25
7 UK	19 (40%)	28	37
8 UK	15 (58%)	11	24
TOTAL (427):	128 (54%)	108	191

☞ Compliance or non-compliance cannot be observed in the control group

ODIN (6-month outcome data)

Follow-up rates depend on Compliance and on Centre

Centre	No. Observations (%)		
	C	NC	Control
1	6 (100%)	2 (22%)	12 (52%)
2	12 (100%)	3 (43%)	7 (64%)
3	17 (100%)	2 (33%)	17 (71%)
4	18 (90%)	6 (75%)	20 (91%)
5	20 (91%)	11 (55%)	17 (68%)
6	17 (100%)	15 (79%)	18 (72%)
7	15 (79%)	16 (57%)	31 (84%)
8	13 (87%)	4 (36%)	18 (75%)
TOTAL:	118 (92%)	59 (55%)	140 (73%)

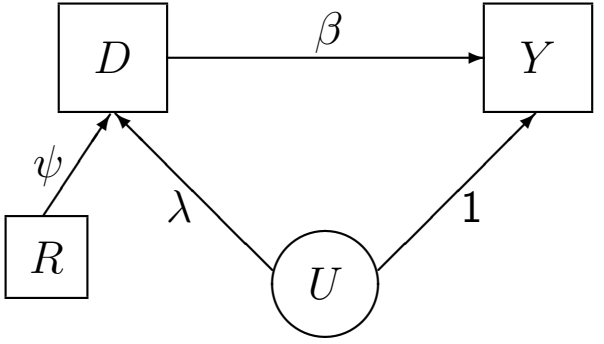
ODIN data (ignoring Centre effects)

	No. at baseline	No. (%) at 6 months	BDI mean
Treatment Group			
Non-Compliers	108	59 (55%)	13.22
Compliers	128	118 (92%)	13.32
Controls	191	140 (73%)	15.16
TOTAL	427	317	

IV modelling with gllamm

Endogenous treatment as a factor model:

D causes Y, with unmeasured confounder U
and instrumental variable R



U is a random effect/latent variable with factor loading λ.

The ODIN study

The data:

R is the randomization indicator (`rgroup`: 0,1).

D is the number of sessions of psychotherapy attended (`sessions`: from 0 to 8).

Y is the BDI score at 6 months (`bdi6`).

U (the unmeasured confounder) is a random effect; it's a latent variable with loading λ.

Remember that there are missing outcome data (assumed to be ignorable)

Model:

$$\begin{aligned} \texttt{bdi6} &= \alpha + \beta\ \texttt{sessions} + U + \varepsilon \\ \texttt{sessions} &= \gamma + \psi\ \texttt{rgroup} + \lambda U + \delta \end{aligned}$$

where $corr(\delta, \varepsilon) = 0$.

Using the two-stage ATR method (Nagelekerke *et al.*) produces $\hat{\beta} = -0.496$ (s.e. 0.312).

Preparing the ODIN data

```
summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rgroup	427	.5526932	.4977989	0	1
sessions	427	2.058548	2.890626	0	8
bdi6	317	14.11356	10.13733	0	46
id	427	214	123.4085	1	427

```
list id rgroup sessions bdi6 in 1/10, clean
```

	id	rgroup	sessions	bdi6
1.	1	1	3	.
2.	2	1	5	0
3.	3	1	6	.
4.	4	0	0	.
5.	5	0	0	.
6.	6	1	0	.
7.	7	1	2	40
8.	8	0	0	18
9.	9	0	0	5
10.	10	1	6	7

Preparing the ODIN data (continued)

```
gen resp1=bdi6
gen resp2=sessions

reshape long resp, i(id) j(type)
```

```
(note: j = 1 2)

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                      427   ->      854
Number of variables                   6   ->        6
j variable (2 values)                     ->     type
xij variables:
                              resp1 resp2   ->     resp
-----------------------------------------------------------------------------
```

```
tab type, gen(d)
```

type	Freq.	Percent	Cum.
1	427	50.00	50.00
2	427	50.00	100.00
Total	854	100.00	

Preparing the ODIN data (continued)

```
list id rgroup type d1 d2 resp in 1/20, clean
```

	id	rgroup	type	d1	d2	resp
1.	1	1	1	1	0	.
2.	1	1	2	0	1	3
3.	2	1	1	1	0	0
4.	2	1	2	0	1	5
5.	3	1	1	1	0	.
6.	3	1	2	0	1	6
7.	4	0	1	1	0	.
8.	4	0	2	0	1	0
9.	5	0	1	1	0	.
10.	5	0	2	0	1	0
11.	6	1	1	1	0	.
12.	6	1	2	0	1	0
13.	7	1	1	1	0	40
14.	7	1	2	0	1	2
15.	8	0	1	1	0	18
16.	8	0	2	0	1	0
17.	9	0	1	1	0	5
18.	9	0	2	0	1	0
19.	10	1	1	1	0	7
20.	10	1	2	0	1	6

Preparing the ODIN data (continued)

```
gen d1_sessions=d1*sessions

gen d2_rgroup=d2*rgroup

eq fac: d1 d2

gllamm resp d1_sessions d1 d2 d2_rgroup, nocons i(id) /*
    */ family(gauss gauss) link(identity identity) fv(type) /*
    */ lv(type) eq(fac) adapt nip(15) trace
```

The gllamm command

```
eq fac: d1 d2
gllamm resp d1_sessions d1 d2 d2_rgroup, nocons i(id) family(gauss gauss) /*
    */ link(identity identity) fv(type) lv(type) eq(fac) adapt nip(15) trace
```

Explanation:

The fixed effects are `d1`, `d1_sessions`, `d2`, and `d2_rgroup`. The random effect (U) is `fac` loading from `d1` and `d2` (the binary indicators for Y and D, respectively).

`nocons`	suppresses the intercept term (represented, instead, by the effects for `d1` and `d2`)
`i(id)`	identifies the participants (level 2 units)
`family(gauss gauss)`	probability distributions for the two outcomes
`link(identity identity)`	link functions for the two outcomes
`fv(type)`	variable whose values indicate which family applies to which observation
`lv(type)`	variable whose values indicate which link function applies to which observation
`eq(fac)`	equation for the latent variable
`adapt nip(15)`	specification for adaptive quadrature

The gllamm output (final part only)

```
...
number of level 1 units = 744
number of level 2 units = 427

gllamm model

log likelihood = -2127.6743
```

resp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1_sessions	-.4958635	.3112457	-1.59	0.111	-1.105894	.1141668
d1	15.15714	.8550292	17.73	0.000	13.48132	16.83297
d2	2.44e-09	.1602771	0.00	1.000	-.3141374	.3141374
d2_rgroup	3.724576	.2155904	17.28	0.000	3.302027	4.147126

```
Variance at level 1

  4.853494 (.34316457)

Variances and covariances of random effects

***level 2 (id)

    var(1): 97.779296 (8.3379229)

    loadings for random effect 1
    d1: 1 (fixed)
    d2: .02329433 (.02173818)
```

gllamm with binary endogenous treatment effects

```
eq fac: d1 d2

gllamm resp d1_treat d1 d2 d2_rgroup, nocons i(id) family(gauss binom) /*
    */ link(identity probit) fv(type) lv(type) eq(fac) adapt nip(15) trace
```

Differences from the previous run:

- Replace `d1_sessions` with corresponding `d1_treat`
- `family(gauss binom)`
- `link (identity probit)`

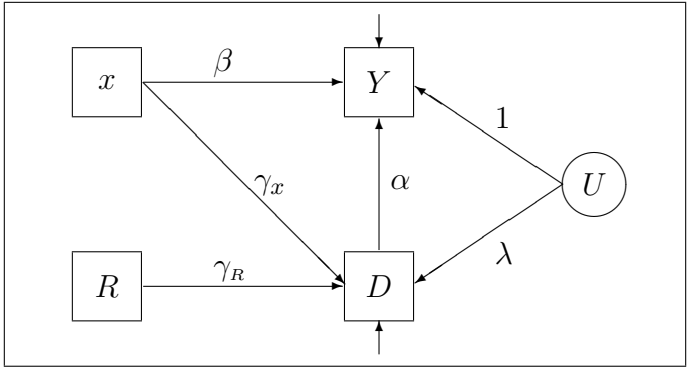
Binary endogenous treatment model: gllamm output

```
...
number of level 1 units = 744
number of level 2 units = 427

log likelihood = -1344.6925
```

resp	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1_treat	-4.259795	2.458733	-1.73	0.083	-9.078823	.5592327
d1	15.36503	.9200239	16.70	0.000	13.56182	17.16824
d2	-16.97098	419.7303	-0.04	0.968	-839.6273	805.6854
d2_rgroup	17.13592	419.732	0.04	0.967	-805.5237	839.7955

```
Variance at level 1
------------------------------------------------------------------------------
  89.246447 (133.98532)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)

    var(1): 15.143656 (134.2019)

    loadings for random effect 1
    d1: 1 (fixed)
    d2: .31621095 (4.8864784)
```

Generalised IV factor model



with a model for Y from the GLM family

$$E(Y_j \mid D_j, x_j, U_j) = g_Y^{-1}(\alpha D_j + \beta x_j + U_j)$$

and similarly for D

$$E(D_j \mid R_j, x_j, U_j) = g_D^{-1}(\gamma_R R_j + \gamma_x x_j + \lambda U_j)$$

where g_Y^{-1} and g_D^{-1} are inverse link functions.

Estimation for non-identity link functions

For g_Y and g_D identity links we have a standard instrumental variable model for the treatment effect α. While incorrect choice of g_D does not lead to inconsistent estimates of the treatment effect α, this is not the case for incorrect choice of g_Y; see e.g. Ten Have *et al.* (2003).

Estimation of models with non-identity links is more complicated. The Stata routine `gllamm` allows an estimation of these models for any appropriate choice of the link function by the explicit integration over the distribution of U using Gaussian, adaptive or non-parametric methods.

Physician advice and drinking example

Kenkel and Terza (2001) analysed 2467 currently drinking males with hypertension.

Data description

- Data from the 1990 National Health Interview Survey.
- Count of alcohol units in last 2 weeks.
- Three dummy explanatory variables:

  `race` (0 = non-black, 1 = black)

  `educ` (high education; 0 if ≤ 12 years, 1 if > 12 years)

  `advice` (told by physician to drink less; 0 = no, 1 = yes)

- There is no randomization to receive advice – instead three IV's are selected on theoretical grounds, i.e.

  `hlthins` (covered by health insurance; 0 = no, 1 = yes)

  `regmed` (registered source of medical care; 0 = no, 1 = yes)

  `heart` (heart condition; 0 = no, 1 = yes)

Physician advice and drinking example

Parameter	Poisson Est	Poisson (SE)	Overdisp. Poisson Est	Overdisp. Poisson (SE)	Probit Est	Probit (SE)	Endog. Treatment Est	Endog. Treatment (SE)
Fixed part								
Drinking model								
α [advice]	**0.47**	(0.01)	**0.59**	(0.08)			**−2.42**	(0.23)
β_0 [cons]	2.65	(0.01)	1.43	(0.06)			2.32	(0.09)
β_1 [hieduc]	−0.18	(0.01)	0.02	(0.07)			−0.29	(0.10)
β_2 [black]	−0.31	(0.02)	−0.29	(0.11)			0.20	(0.11)
Advice model								
γ_0 [cons]					−0.48	(0.08)	−1.13	(0.16)
γ_1 [hieduc]					−0.25	(0.06)	−0.40	(0.10)
γ_2 [black]					0.30	(0.08)	0.60	(0.15)
γ_3 [hlthins]					−0.27	(0.07)	−0.33	(0.10)
γ_4 [regmed]					0.18	(0.07)	0.39	(0.10)
γ_5 [heart]					0.17	(0.08)	0.51	(0.11)
Random part								
Variance								
ψ			2.90	(0.11)			2.50	(0.69)
Loading								
λ							1.43	(0.15)
Log likelihood	−32939.15		−8857.85		−1419.90		−10254.02	

CACE: Complier Average Causal Effect

Two Types of Patient

1. **Complier** – Accepts allocation

   ☞ Can be identified (C) in the Treatment Group, but hidden or latent in the Controls

2. **Non-Complier** – Would never receive therapy, whatever the allocation

   ☞ Can be identified (NC) in the Treatment Group, but hidden or latent in the Controls

CACE Estimation

Assumption 1: Randomization

Randomization ensures that, on average, **the proportions of Compliers and Non-Compliers are the same in the two arms of the trial**.

Therefore these proportions can be estimated from the observed proportions in the Treatment Group.

CACE Estimation

Assumption 2: Exclusion Restriction

For **Non-Compliers**, the outcome is the same in the two arms of the trial.

That is, **the offer of treatment, in itself, does not influence outcome**.

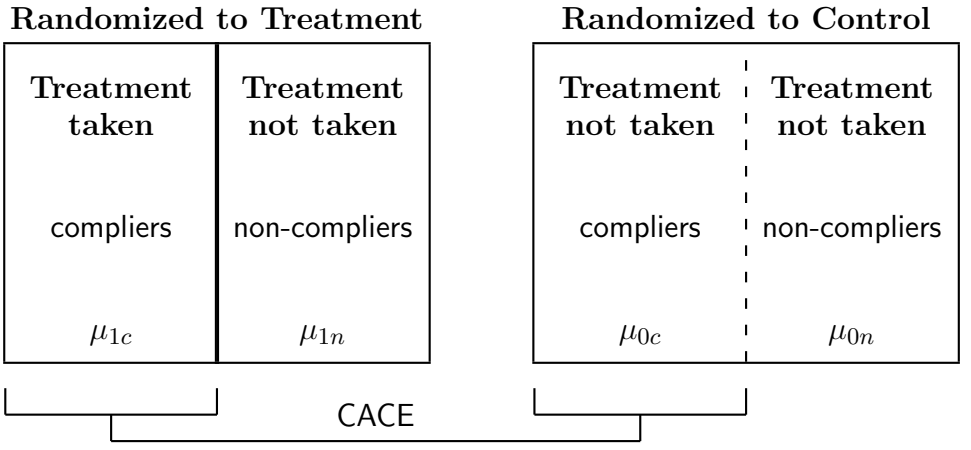
CACE Estimation

Assumption 3: Ignorable Missing Data Mechanism

Data Missing At Random (MAR) - i.e. Ignorable

Given observed Compliance Status (Complier, Non-Complier or Control), outcome is independent of whether it is actually observed or not.

Complier Average Causal Effect



- CACE is treatment effect for compliers

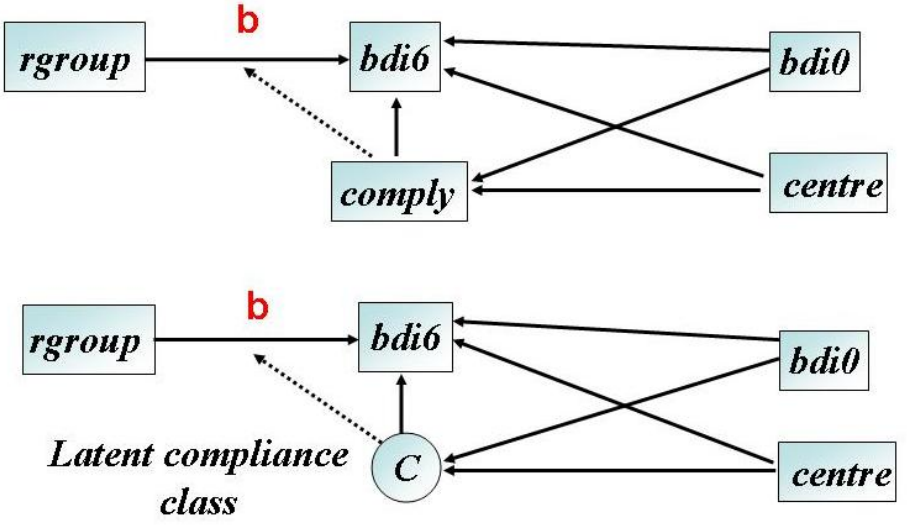
$$\delta_c = \mu_{1c} - \mu_{0c},$$

  μ_{1c} and μ_{0c} mean outcomes of compliers in treatment and control groups

- Exclusion restriction: mean outcome same among non-compliers in both groups

$$\mu_{1n} = \mu_{0n}$$

Path diagram for CACE latent class model



b: in class 1 (non-compliers) this path is fixed at 0

in class 2 (compliers) this path is free

Outcome model

- r_j is dummy for being randomized to treatment versus control
- c_j is dummy for compliers versus non-compliers
- Model for outcome if compliance were known for everyone:

$$y_j = \beta_0 + \beta_1 c_j(1 - r_j) + \beta_2 c_j r_j + \epsilon_j,$$

  - c_j observed only if $r_j = 1$, i.e. in third term
  - c_j in second term never observed: discrete latent variable $\eta_j = e_1, e_2$, where $e_1 = 1$, $e_2 = 0$:

$$\text{Depression model: } \; y_j = \beta_0 + \beta_1 \eta_j(1 - r_j) + \beta_2 c_j r_j + \epsilon_j$$

  - CACE:

$$\mu_{1n} = \mu_{0n} = \beta_0, \quad \mu_{0c} = \beta_0 + \beta_1, \quad \mu_{1c} = \beta_0 + \beta_2$$
$$\Longrightarrow \delta_c = \beta_2 - \beta_1$$

Compliance model

- Probability of being complier same in treatment and control groups (due to randomisation)

$$\Pr(c_j = 1 \mid r_j = 1) = \Pr(c_j = 1 \mid r_j = 0) = \Pr(\eta_j = e_1) = \pi_1$$

- Without covariates for compliance

$$\text{Compliance model: } \operatorname{logit}[\Pr(c_j = 1)] = \varrho = \operatorname{logit}(\pi_1)$$

CACE in gllamm (continued)

Response model: $\nu_{ij} = \beta_0 d_{i1} + \beta_1 \eta_j (1 - r_j) d_{i1} + \beta_2 c_j r_j d_{i1} + \varrho d_{i2}$

Structural model: $\text{logit}[\pi_1] = \varrho.$

- Interactions and equations:

```
gen c_r_d1 = c*r*d1          /* c_j r_j d_i1 */
gen nr_d1 = (1-r)*d1         /* (1 - r_j) d_i1 */
eq load: nr_d1               /* for β_1 (1 - r_j) d_j1 */
```

- Constraints:

```
cons def 1 [z2_1_1]nr_d1 = 1          /* e_1 = 1 */
cons def 2 [z2_1_2]nr_d1 = 0          /* e_2 = 0 */
cons def 3 [p2_1]_cons = [y]d2        /* constraint for ϱ */
```

- gllamm command:

```
gllamm y d1 c_r_d1 d2, i(id) eqs(load) l(ident logit) /*
  */ f(gauss binom) lv(var) fv(var) ip(fn) nip(2)     /*
  */ constr(1/3) frload(1) nocons    /* β_1 is 'freed' by frload(1) */
```

gllamm command line and part output

```
eq load: nr_d1

cons def 1 [z2_1_1]nr_d1=1
cons def 2 [z2_1_2]nr_d1=0
cons def 3 [p2_1]_cons=[y]d2

gllamm y d1 c_r_d1 d2, i(id) eqs(load) l(ident logit) f(gauss binom) /*
   */ lv(var) fv(var) ip(fn) nip(2) const(1/3) frload(1) nocons
```

```
...
number of level 1 units = 553
number of level 2 units = 376
log likelihood = -1344.824804098342
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1	13.07684	1.282891	10.19	0.000	10.56242	15.59126
c_r_d1	.2451937	1.578327	0.16	0.877	-2.848271	3.338659
d2	.1664994	.1311042	1.27	0.204	-.09046	.4234588

ODIN CACE Estimate assuming MAR (no covariates)

```
Probabilities and locations of random effects
------------------------------------------------------------------------------
***level 2 (id)
    loc1: 1, 0
  var(1): .24827535

    loadings for random effect 1
    nr_d1: 3.9636215 (2.8651214)
    prob: 0.5415, 0.4585
------------------------------------------------------------------------------
```

```
lincom [y]c_r_d1-[id1_1l]nr_d1
```

```
 ( 1)  [y]c_r_d1 - [id1_1l]nr_d1 = 0
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
(1)	-3.718428	2.151174	-1.73	0.084	-7.934651	.4977952

gllamm MAR setup with covariates


```
eq p: c1 c2 c3 c4 c6 c7 c8 bdi0
eq load: nr_d1
cons def 1 [z2_1_1]nr_d1=1
cons def 2 [z2_1_2]nr_d1=0
cons def 3 [p2_1]_cons=[y]d2
cons def 4 [p2_1]c1=[y]c1_d2
cons def 5 [p2_1]c2=[y]c2_d2
cons def 6 [p2_1]c3=[y]c3_d2
cons def 7 [p2_1]c4=[y]c4_d2
cons def 8 [p2_1]c6=[y]c6_d2
cons def 9 [p2_1]c7=[y]c7_d2
cons def 10 [p2_1]c8=[y]c8_d2
cons def 11 [p2_1]bdi0=0

gllamm y d1 c_r_d1 d2 /*
  */ c1_d1 c2_d1 c3_d1 c4_d1 c6_d1 c7_d1 c8_d1 bdi0_d1/*
  */ c1_d2 c2_d2 c3_d2 c4_d2 c6_d2 c7_d2 c8_d2 bdi0_d2 ,/*
  */ i(id) eqs(load) peqs(p) l(ident logit) f(gauss binom) lv(var) /*
  */ fv(var) ip(fn) nip(2) const(1/11) frload(1) from(a) skip nocons

lincom [y]c_r_d1-[id1_1l]nr_d1
```

ODIN CACE Estimate assuming MAR (no covariates)

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1	.1421738	1.863593	0.08	0.939	-3.510401	3.794748
c_r_d1	1.633098	1.222449	1.34	0.182	-.7628577	4.029053
d2	-.2732483	.4660384	-0.59	0.558	-1.186667	.6401702
c1_d1	-4.908202	2.239007	-2.19	0.028	-9.296575	-.5198285
c2_d1	-4.246159	2.133657	-1.99	0.047	-8.42805	-.0642679
c3_d1	-2.941429	1.893257	-1.55	0.120	-6.652146	.7692869
c4_d1	-5.314535	1.759019	-3.02	0.003	-8.76215	-1.86692
c6_d1	4.058566	1.66644	2.44	0.015	.7924048	7.324728
c7_d1	4.606668	1.623953	2.84	0.005	1.423778	7.789558
c8_d1	1.544861	1.889072	0.82	0.413	-2.157652	5.247374
bdi0_d1	.5235755	.0601072	8.71	0.000	.4057676	.6413833
c1_d2	-.5903473	.5953044	-0.99	0.321	-1.757123	.5764279
c2_d2	.6116215	.5606669	1.09	0.275	-.4872655	1.710509
c3_d2	1.09433	.5549307	1.97	0.049	.006686	2.181974
c4_d2	.8314038	.5139704	1.62	0.106	-.1759596	1.838767
c6_d2	-.2037023	.444397	-0.46	0.647	-1.074704	.6672999
c7_d2	-.3702143	.4211024	-0.88	0.379	-1.19556	.4551312
c8_d2	.1340363	.4932205	0.27	0.786	-.8326582	1.100731
bdi0_d2	.0140192	.0160298	0.87	0.382	-.0173987	.0454371

CACE Estimate (cont'd)


```
Probabilities and locations of random effects
------------------------------------------------------------------------------
***level 2 (id)
    loc1: 1, 0
  var(1): .24539092
    loadings for random effect 1
    nr_d1: 6.8980394 (2.0763784)

    prob: 0.4321, 0.5679

    log odds parameters
    class 1
    c1: -.59034731 (.59530442)
    c2: .61162153 (.56066694)
    c3: 1.0943302 (.55493073)
    c4: .83140379 (.51397035)
    c6: -.20370229 (.44439705)
    c7: -.37021427 (.42110239)
    c8: .13403632 (.49322054)
    bdi0: .01401919 (.01602984)
    _cons: -.27324832 (.46603843)
------------------------------------------------------------------------------
```

```
lincom [y]c_r_d1 - [id1_1l]nr_d1
```

```
( 1)  [y]c_r_d1 - [id1_1l]nr_d1 = 0
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
(1)	-5.264942	1.646564	-3.20	0.001	-8.492148	-2.037736

CACE Estimation

Alternative Assumption 3: Latent Ignorability

Given both Treatment Arm (Z=0 or 1) and Compliance Status (Complier, Non-Complier), outcome is independent of whether it is actually observed or missing.

The 'Latent' in 'Latent Ignorability' comes from the fact that we cannot observe compliance status completely.

CACE Estimation

Alternative Assumption 2: Compound Exclusion Restriction

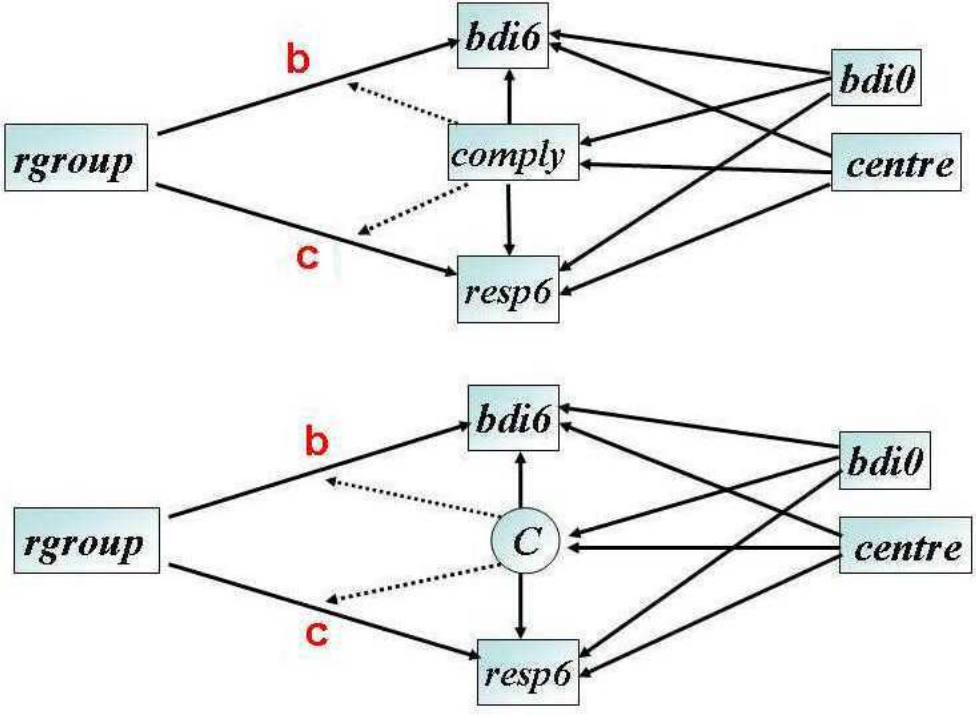
For Non-Compliers, the drop-out rate is the same in the two arms of the trial.

That is, the offer of treatment, in itself, does not influence loss to follow-up.

For Non-Compliers, the outcome is the same in the two arms of the trial.

That is, the offer of treatment, in itself, does not influence outcome.

CACE model with latent ignorable



b, c: in class 1 (non-compliers) this path is fixed at 0
in class 2 (compliers) this path is free

gllamm setup for CACE estimate assuming LI (no covariates)

```
                                *!Addition of factor loading from latent
eq load: nr_d1 nr_d3            *!compliance class to missingess indicator

cons def 1 [z2_1_1]nr_d1=1
cons def 2 [z2_1_2]nr_d1=0
cons def 3 [p2_1]_cons=[y]d2

gllamm y d1 c_r_d1 d2 d3 c_r_d3, i(id) eqs(load) l(iden logit logit) /*
   */f(gauss binom binom) lv(var) fv(var) ip(fn) nip(2) const(1/3) /*
   */ frload(1) nocons
```

ODIN CACE Estimate assuming LI (no covariates)

```
number of level 1 units = 980
number of level 2 units = 427
log likelihood = -1565.20793100089
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1	13.16542	1.284933	10.25	0.000	10.647	15.68384
c_r_d1	.156616	1.58229	0.10	0.921	-2.944615	3.257847
d2	.1687961	.1307435	1.29	0.197	-.0874565	.4250486
d3	.1877536	.1934712	0.97	0.332	-.1914431	.5669502
c_r_d3	2.280347	.3819762	5.97	0.000	1.531687	3.029006

ODIN CACE Estimate assuming LI (no covariates)

```
Probabilities and locations of random effects
------------------------------------------------------------------------------
***level 2 (id)
    loc1: 1, 0
  var(1): .24822767

    loadings for random effect 1
    nr_d1: 3.0631954 (2.3352733)
    nr_d3: 1.902311 (.88119373)

    prob: 0.5421, 0.4579
------------------------------------------------------------------------------
```

```
. lincom [y]c_r_d1 - [id1_1l]nr_d1

 ( 1)  [y]c_r_d1 - [id1_1l]nr_d1 = 0

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -2.906579    1.72964    -1.68   0.093    -6.296611     .4834523
------------------------------------------------------------------------------
```

References and select bibliography

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* **91**, 444-472.

Boxall, P. and Adamowicz, W. (2002). Understanding heterogeneous preferences in random utility models: a Latent class approach. *Environmental and Resource Economics* **23**(4), 421-446.

Dawid, A.P. (2000). Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association* **95**, 407-448.

Dunn, G. Maracy, M., Dowrick, C., *et al.* (2003). Estimating psychological treatment effects from an RCT with both non-compliance and loss to follow-up: the ODIN Trial. *British Journal of Psychiatry* **183**, 323-331.

Greene, W.H. and Hensher, D.A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B* **37**(8), 681-698.

Imbens, G.W. and Rubin, D.B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies* **64**, 555-574.

Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of intervention with noncompliance. *Statistics in Medicine* **21**, 3161-3181.

Kenkel, D.S. and Terza, J.V. (2001). The effect of physician advice on alcohol consumption: count regression with an endogenous treatment effect. *Journal of Applied Econometrics* **16**, 165-184.

Little, R. and Yau, L.H.Y. (1998). Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychological Methods* **3**, 147-159.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000). *Stated choice methods: analysis and application*. Cambridge University Press.

Maughan, B., Pickles, A., Rowe, E., Costello, J. and Angold, A. (2000). Developmental trajectories of aggressive and non-aggressive conduct problems. *International Journal of Quantitative Criminology* **16**, 199-221.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Pages 105–142 of:* Zarembka, P. (ed), *Frontiers in Econometrics*, Academic Press, New York.

Muthén, B.O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika* **29**, 81-117. Downloadable from *http://www.statmodel.com/muthen1.pdf*

Rabe-Hesketh, S., Pickles, A. and Taylor, C. (2000). sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin* **53**, 47-57.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* **2**, 1-21.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004). *GLLAMM Manual*. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160 (*http://www.bepress.com/ucbbiostat/paper160*).

Skrondal, A. and Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika* **68**, 267-287.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal and structural equation models*. Chapman & Hall/CRC Press, Boca Raton, FL.

Snidjers, T. and Bosker, R. (1999). *Multilevel analysis.* Sage Publications, London.

Ten Have, T.R., Joffe, M. and Cary, M. (2003). Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine* **22**, 1255-1283.

Vinokur, A.D., Price, R.H. and Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology* **23**, 39-74.

Yau, L.H.Y. and Little, R.J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* **96**, 1232-1243.

Yellott, J. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology* **15**, 109–44.