Generalizing sem in Stata

Jeff Pitblado

Director of Statistical Software

StataCorp LP

2013 Stata Conference

Feature requests

Can **sem** support outcomes that are

- binary?
- counts?
- ordinal?
- categorical?

Is there a command that is like

- **oprobit** for panel data?
- **xtmelogit** but uses probit instead?
- **xtmepoisson** but uses negative binomial instead?

New command

The new **gsem** command

- Think **G**eneralized **S**tructural **E**quations **M**odel
- Inspired by **gllamm** and **sem**
- Documented in [SEM]

Features

- Discrete and continuous outcomes
- Multi-level latent variables
- Maximum likelihood estimator
- Supported by the SEM Builder

Modeling outcomes

Syntax

gsem *y* **<-** *x* **,** **family(***family***)** **link(***link***)**

Supported family and link combinations

	logit	probit	cloglog	log	identity
Bernoulli	x	x	x		
binomial	x	x	x		
ordinal	x	x	x		
multinomial	x				
Poisson				x	
negative binomial				x	
gamma				x	
Gaussian				x	x

Latent variables

Syntax

- Latent variables – **sem** style
  - Capitalized names
  - **latent()** option
- Multi-level latent variables – notice square brackets
  - *Lname*[*level_spec*]
- Random slopes – uses interaction notation
  - *Lname*[*level_spec*]#...

Integration methods

- Mean-variance adaptive quadrature
- Mode-curvature adaptive quadrature
- Gauss-Hermite quadrature
- Laplacian approximation

Maximum likelihood estimator

Model construction

- Outcomes are independent of each other, conditionally on
  - Latent variables
  - Other observed variables, with an exception
    - Linear outcomes (Gaussian/identity) modeled jointly
- Covariance estimation
  - Linear outcome errors
  - Latent variables within a given level

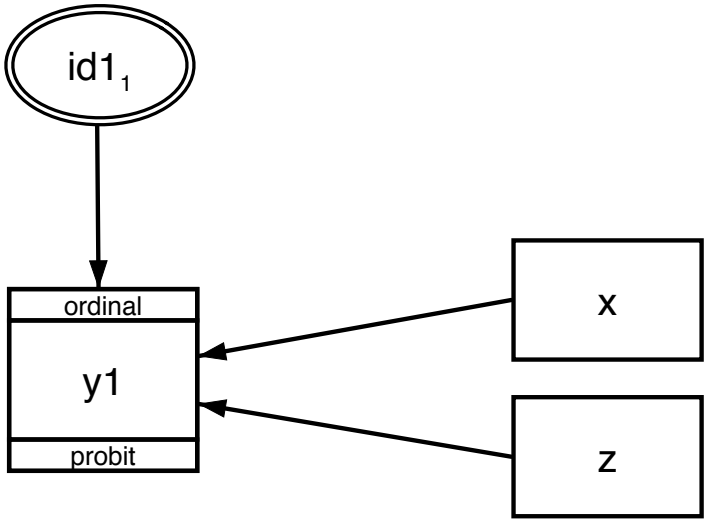
Multivariate outcomes models

- All recursive models are allowed
- Non-recursive models (loops) only allowed for linear outcomes

Stata DEMO

SEM Builder Demonstration

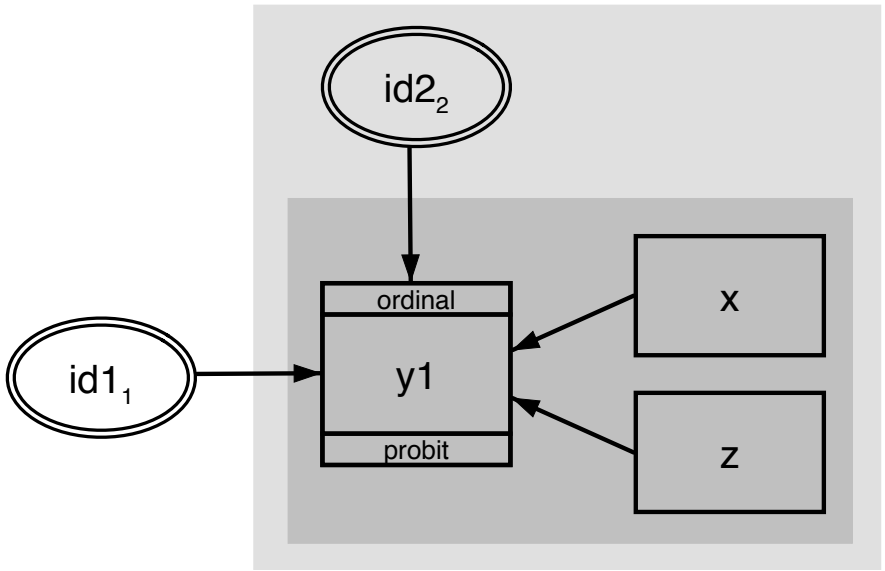
Example: oprobit for panel data




```
gsem y1 <- x z l[id1], family(ordinal) link(probit)

xtoprobit y1 x z
```

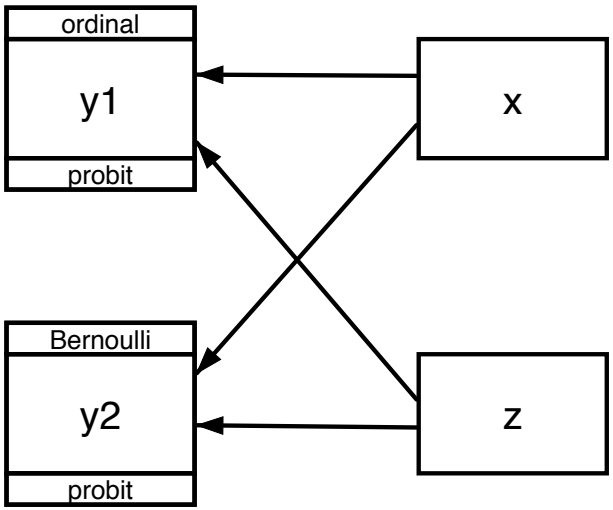
Example: oprobit for multi-level data



gsem y1 <- x z I[id1] I[id1>id2], oprobit

meoprobit y1 x z II id1: II id2:

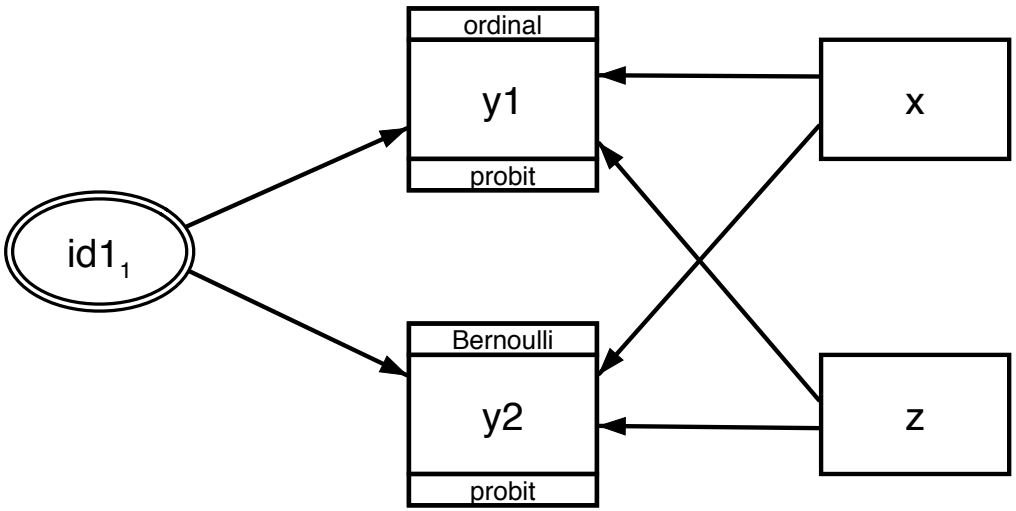
Example: seemingly unrelated estimation




```
gsem (y1 <- x z, oprobit) (y2 <- x z, probit) , vce(robust)
```

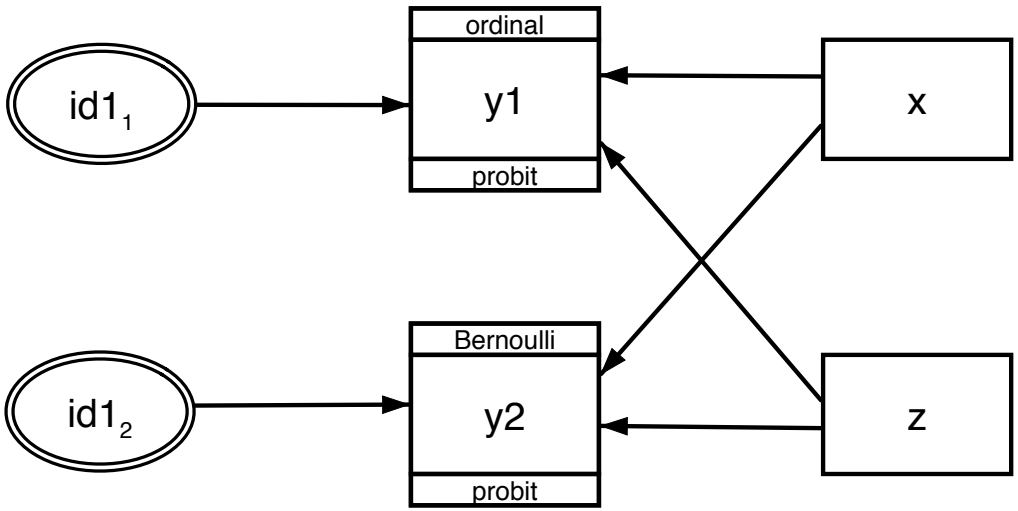
```
oprobit y1 x z
estimates store y1
probit y2 x z
suest y1 .
```

Example: bivariate response with a random effect



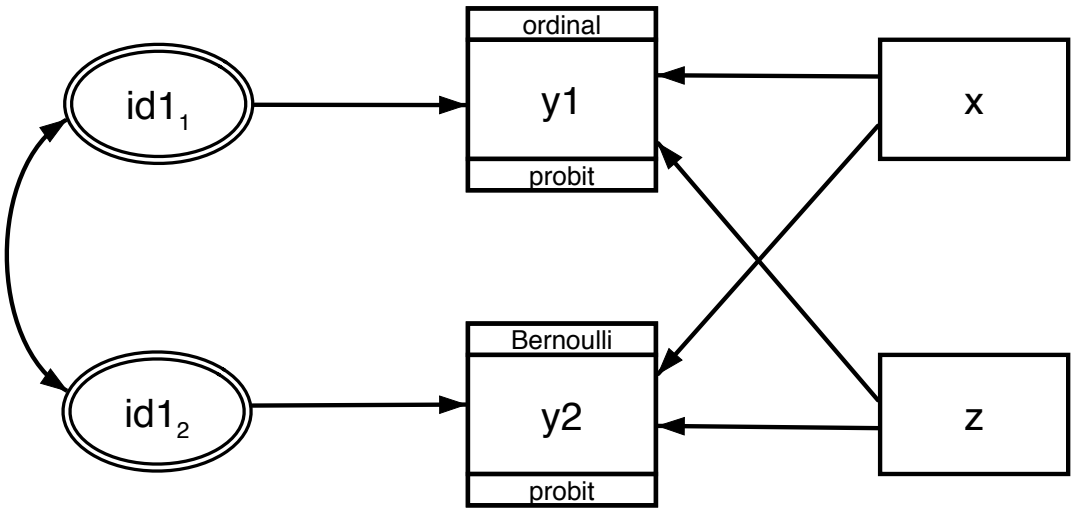
gsem (y1 <- x z I[id1], oprobit) (y2 <- x z I[id1], probit)

Example: bivariate, with 2 independent random effects




```
gsem (y1 <- x z I[id1], oprobit) (y2 <- x z J[id1], probit),
        covariance(I[id1]*J[id1]@0)
```

Example: bivariate, with 2 random effects



gsem (y1 <- x z I[id1], oprobit) (y2 <- x z J[id1], probit)

Conclusion

What's next

- more family/link combinations
- alternatives to quadrature
- improve performance for linear outcomes models
- . . .