Data Analysis and Statistical Software

Generalized Structural Equation Modeling Using Stata

Chuck Huber

StataCorp

Italian Stata Users Group Meeting
November 14-15, 2013

Outline

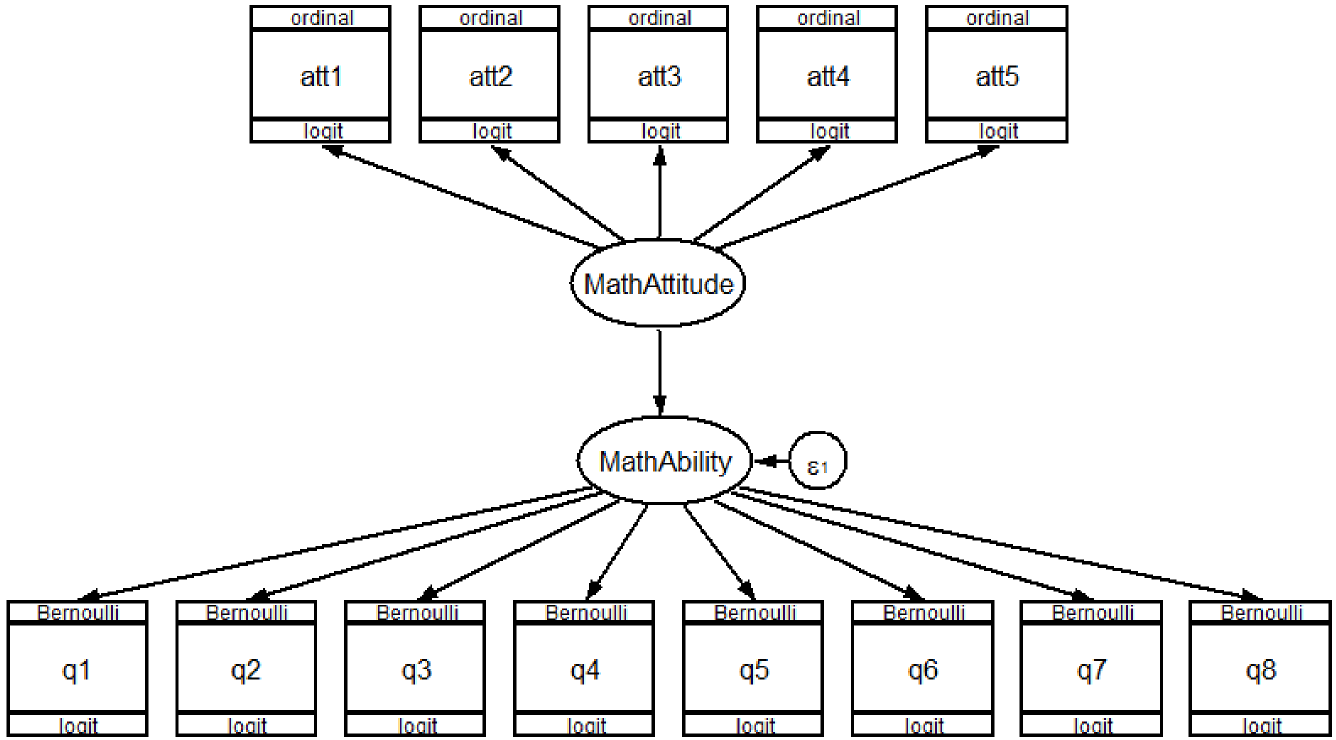
- Introduction to SEM concepts and jargon
- Continuous outcome models using SEM
- Generalized outcome models using GSEM
- Multilevel generalized models using GSEM

What is Structural Equation Modeling?

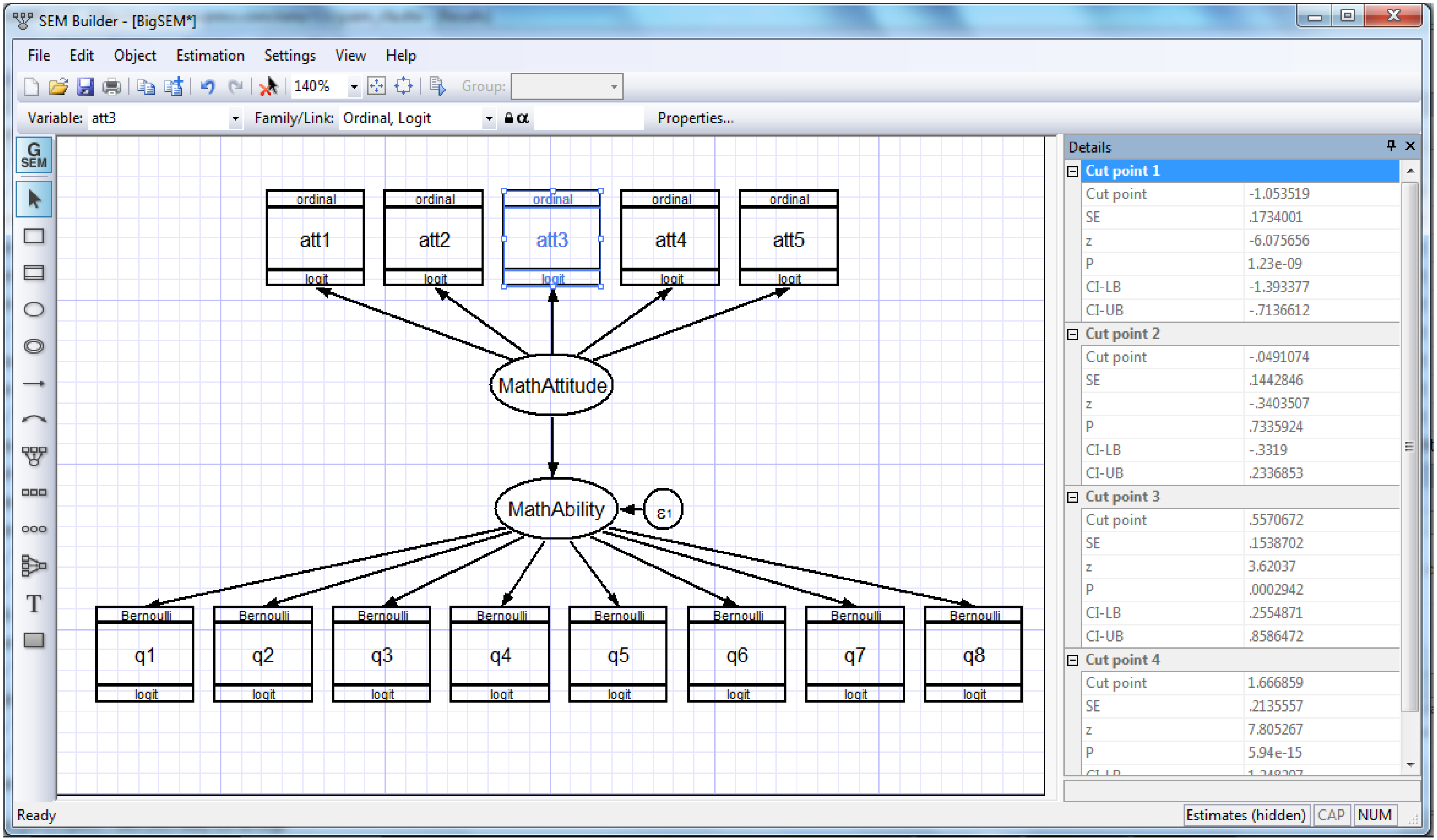
- Structural equation modeling encompasses a broad array of models from linear regression to measurement models to simultaneous equations.
- Structural equation modeling is not just an estimation method for a particular model.
- Structural equation modeling is a way of thinking, a way of writing, and a way of estimating.

-Stata SEM Manual, pg 2

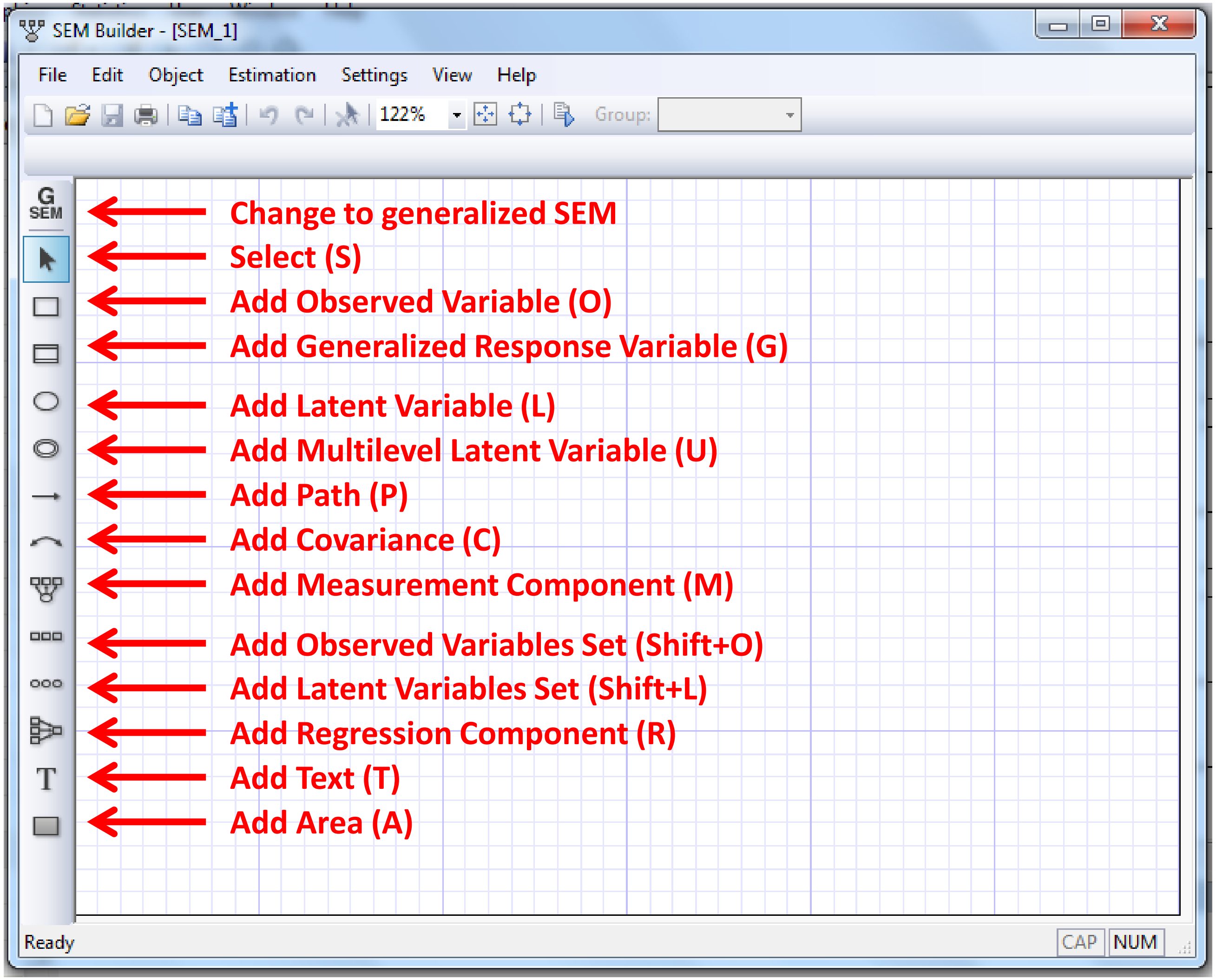
Structural Equation Models are often drawn as **Path Diagrams**:



We can draw path diagrams using Stata's SEM Builder



We can draw path diagrams using Stata's SEM Builder



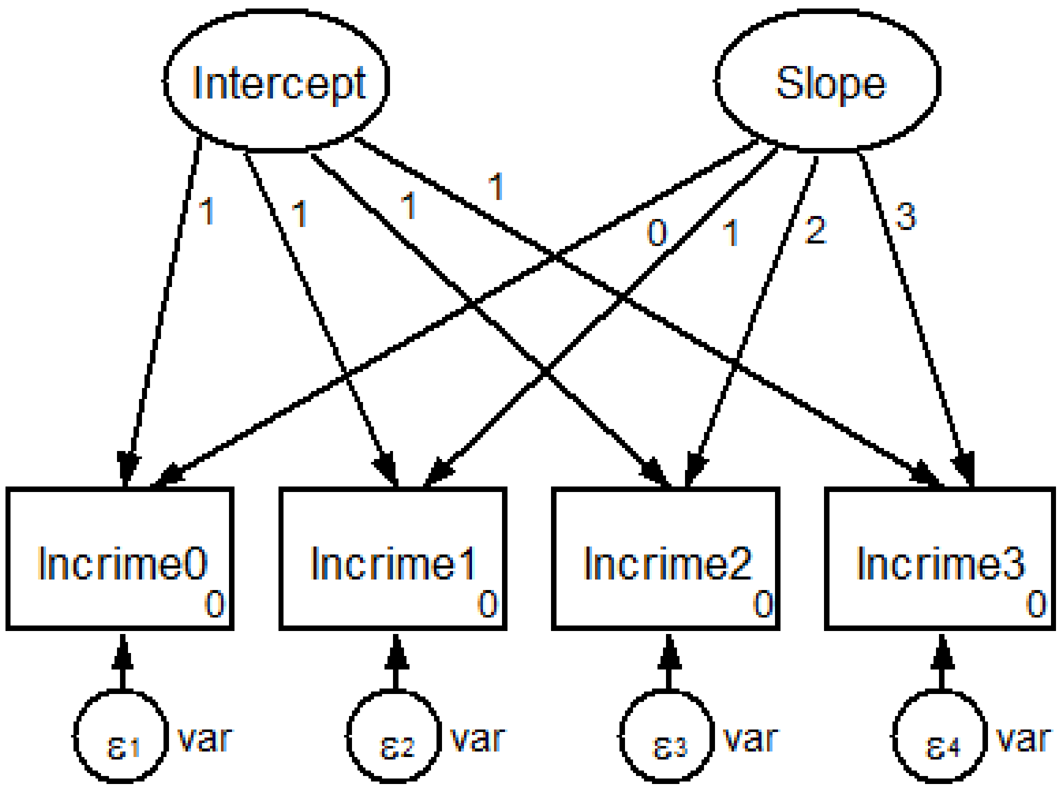
Jargon

- SEM and GSEM
- Observed and Latent variables
- Paths and Covariance
- Endogenous and Exogenous variables
- Recursive and Nonrecursive models

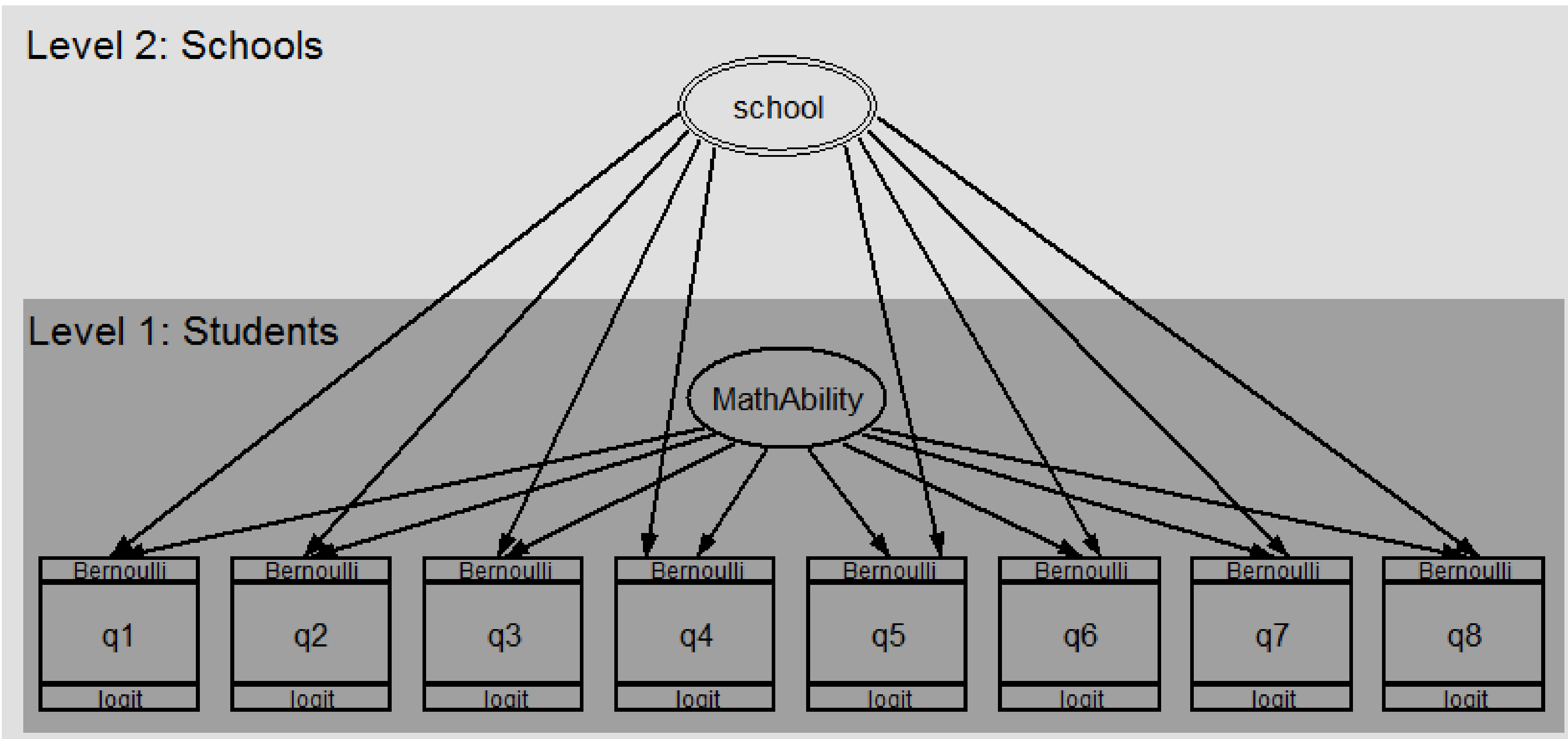
SEM vs GSEM?

- Structural Equation Modeling (SEM)
  - Continuous outcomes
  - Single level data structures
  - Compatible with **`-svy-`**
- Multilevel Generalized Outcomes (GSEM)
  - Generalized responses (binary, ordered, count, etc)
  - Multilevel data structures
  - Can use factor variable notation

Structural Equation Model (SEM)



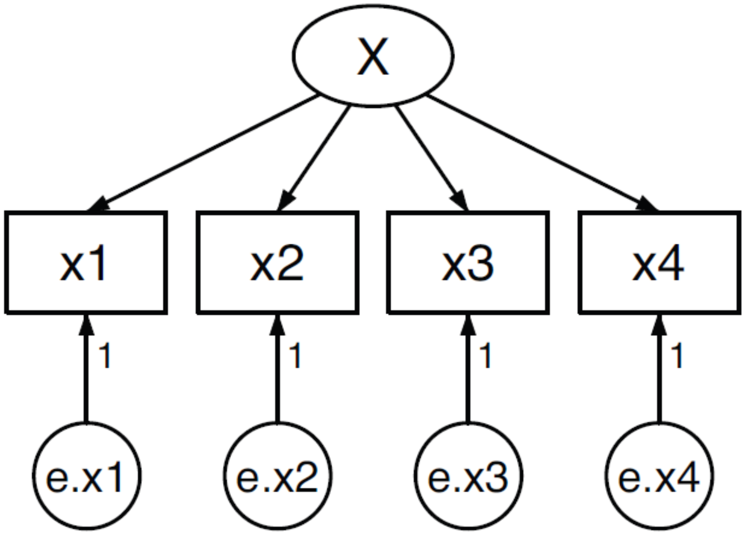
Generalized Structural Equation Model (GSEM)



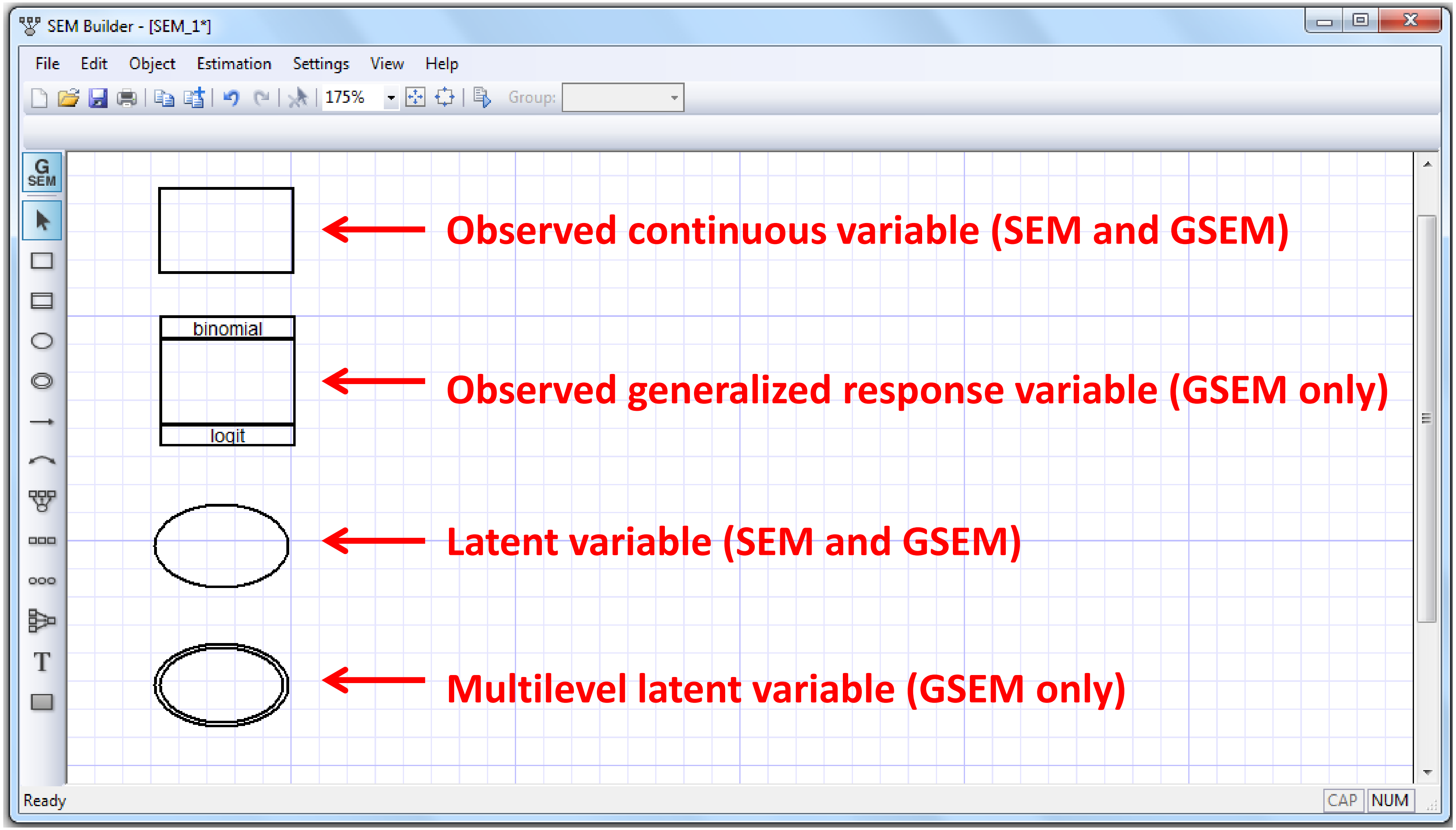
Observed and Latent Variables

- **Observed variables** are variables that are included in our dataset. They are represented by rectangles. The variables x1, x2, x3 and x4 are observed variables in this path diagram.

- **Latent variables** are unobserved variables that we wish we had observed. They can be thought of as a composite score of other variables. They are represented by ovals. The variable X is a latent variable in this path diagram.

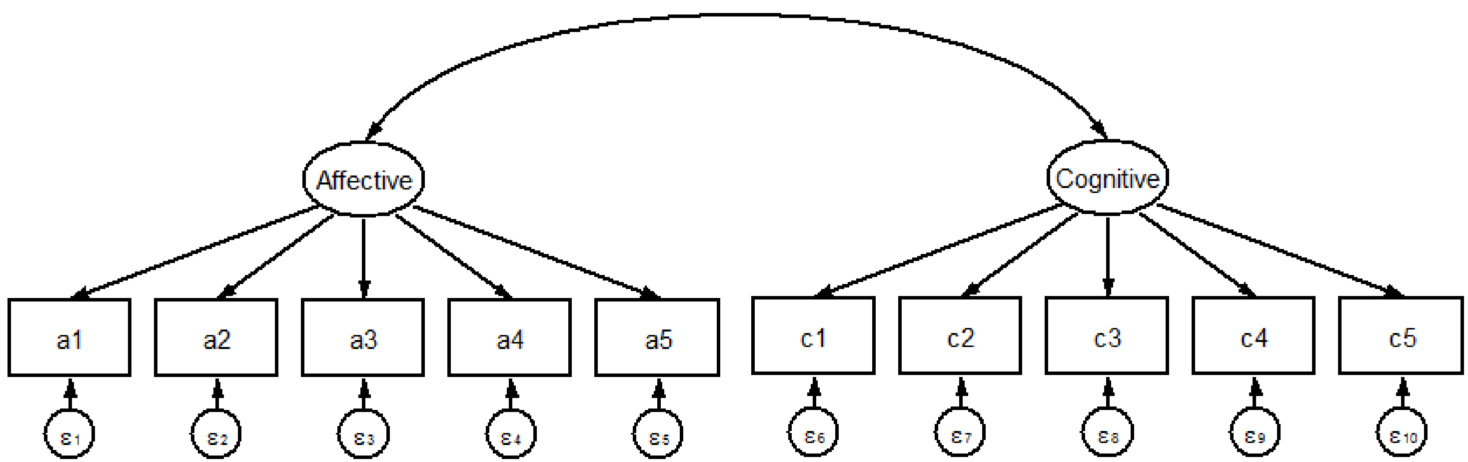


Drawing variables in Stata's SEM Builder



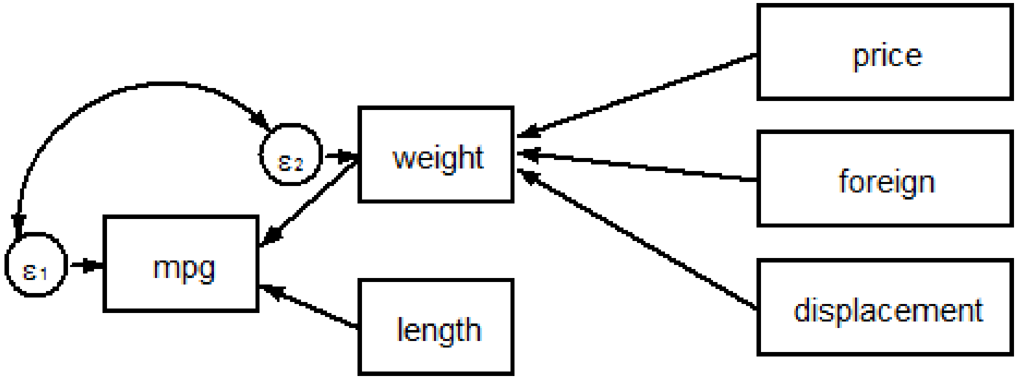
Paths and Covariance

- **Paths** are direct relationships between variables. Estimated path coefficients are analogous to regression coefficients. They are represented by straight arrows.
- **Covariance** specify that two latent variables or error terms covary. They are represented by curved arrows.



Exogenous and Endogenous Variables

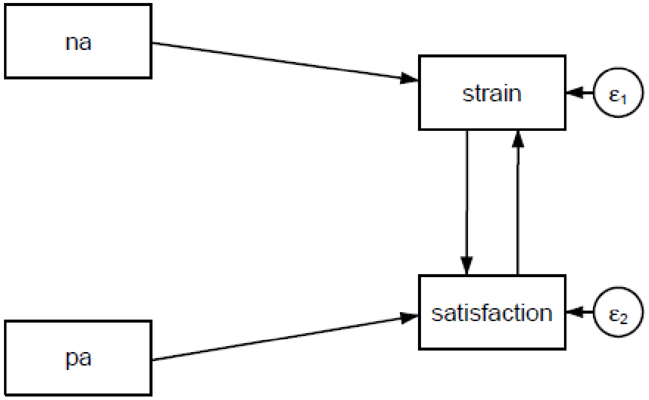
- **Exogenous** variables are determined outside the system of equations. There are no paths pointing to it. The variables `price`, `foreign`, `displacement` and `length` are exogenous.
- **Endogenous** variables are determined by the system of equations. At least one path points to it. The variables `weight` and `mpg` are endogenous.

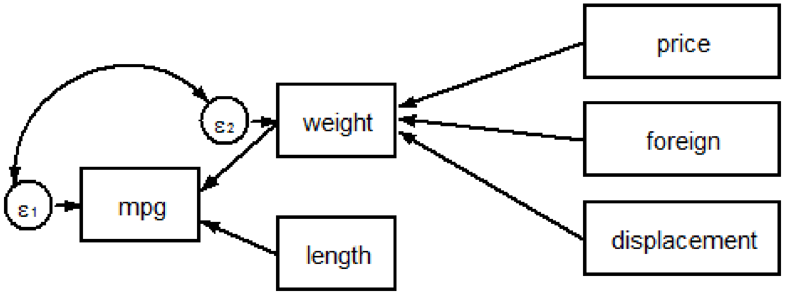


- **Observed Exogenous**: a variable in a dataset that is treated as exogenous in the model
- **Latent Exogenous**: an unobserved variable that is treated as exogenous in the model.
- **Observed Endogenous**: a variable in a dataset that is treated as endogenous in the model
- **Latent Endogenous**: an unobserved variable that is treated as endogenous in the model.

Recursive and Nonrecursive Systems

- **Recursive** models do not have any feedback loops or correlated errors.
- **Nonrecursive** models have feedback loops or correlated errors. These models have paths in both directions between one or more pairs of endogenous variables





Outline

- Introduction to SEM concepts and jargon
- **Continuous outcome models using SEM**
- Generalized outcome models using GSEM
- Multilevel generalized models using GSEM

Continuous outcome models using SEM

- Sample means
- Pearson correlation coefficient
- Student's t-test
- Linear regression
- Multivariate linear regression
- Seemingly unrelated regression
- Three-stage least squares

Continuous outcome models using SEM

```
. sysuse auto
```

variable name	storage type	display format	value label	variable label
make	str18	%-18s		Make and Model
price	int	%8.0gc		Price
mpg	int	%8.0g		Mileage (mpg)
rep78	int	%8.0g		Repair Record 1978
headroom	float	%6.1f		Headroom (in.)
trunk	int	%8.0g		Trunk space (cu. ft.)
weight	int	%8.0gc		Weight (lbs.)
length	int	%8.0g		Length (in.)
turn	int	%8.0g		Turn Circle (ft.)
displacement	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type

Sample Mean Path Diagram

Sample Mean Syntax

Syntax using `means`:

```
mean mpg
```

Syntax using `sem`:

```
sem mpg
```

Sample Mean Results

Results using `means`:

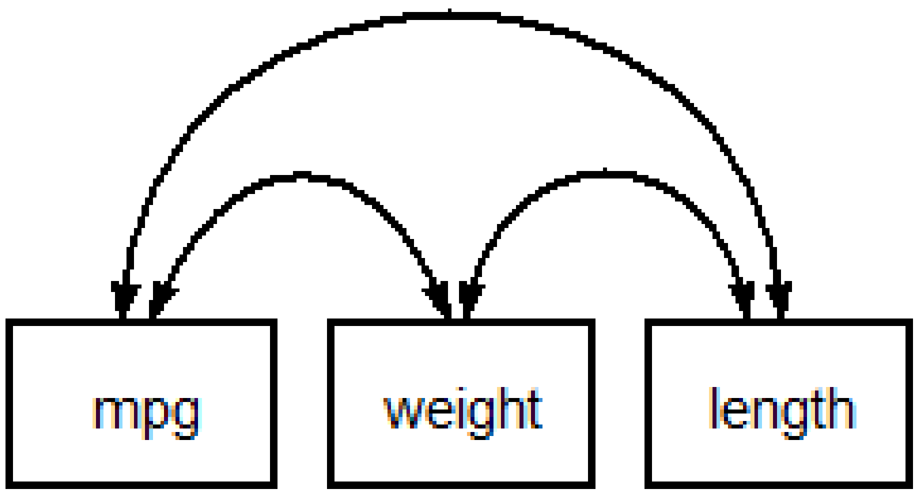
Mean estimation   Number of obs = 74

	Mean	Std. Err.	[95% Conf.	Interval]
mpg	21.2973	.6725511	19.9569	22.63769

Results using `sem`:

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
mean(mpg)	21.2973	.6679914	31.88	0.000	19.98806	22.60654
var(mpg)	33.01972	5.428409			23.92416	45.57326

Correlation Path Diagram



Correlation Syntax

Syntax using `correlate`:

```
correlate mpg weight length
```

Syntax using `sem`:

```
sem mpg weight length, standardized
```

Correlation Results

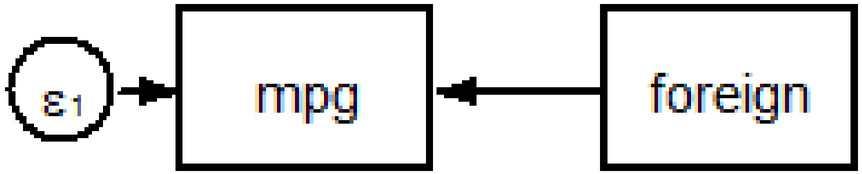
Results using `correlate`:

	mpg	weight	length
mpg	1.0000		
weight	-0.8072	1.0000	
length	-0.7958	0.9460	1.0000

Results using `sem`:

Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
mean(mpg)	3.706276	.3260791	11.37	0.000	3.067173	4.34538
mean(weight)	3.9116	.3419006	11.44	0.000	3.241487	4.581713
mean(length)	8.497816	.7081231	12.00	0.000	7.10992	9.885712
var(mpg)	1	.			.	.
var(weight)	1	.			.	.
var(length)	1	.			.	.
cov(mpg,weight)	-.8071749	.0405087	-19.93	0.000	-.8865704	-.7277793
cov(mpg,length)	-.7957794	.0426321	-18.67	0.000	-.8793368	-.7122221
cov(weight, length)	.9460086	.0122139	77.45	0.000	.9220699	.9699474

Student's t-test Path Diagram



Student's t-test Syntax

Syntax using `ttest`:

```
ttest mpg, by(foreign)
```

Syntax using `sem`:

```
sem mpg <- foreign
```

Student's t-test Results

Results using `ttest`:

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Domestic	52	19.82692	.657777	4.743297	18.50638	21.14747
Foreign	22	24.77273	1.40951	6.611187	21.84149	27.70396
combined	74	21.2973	.6725511	5.785503	19.9569	22.63769
diff		-4.945804	1.362162		-7.661225	-2.230384

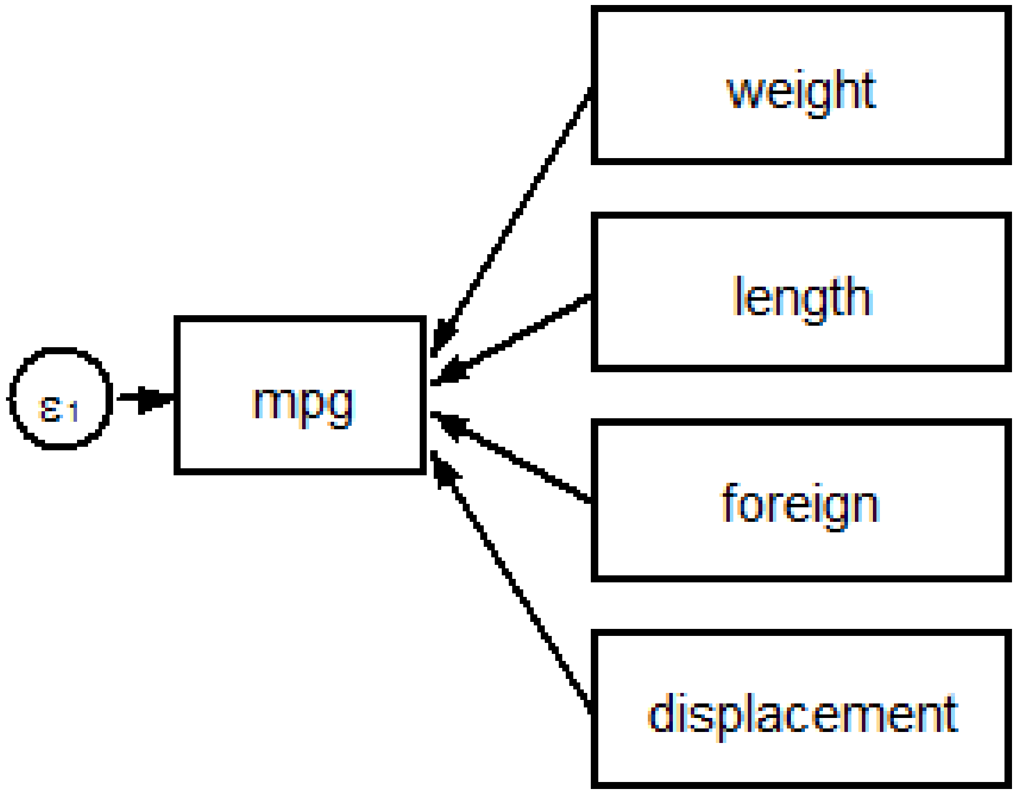
```
    diff = mean(Domestic) - mean(Foreign)                      t =  -3.6308
Ho: diff = 0                                  degrees of freedom =       72

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0003         Pr(|T| > |t|) = 0.0005          Pr(T > t) = 0.9997
```

Results using `sem`:

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
mpg <-						
foreign	4.945804	1.343628	3.68	0.000	2.312341	7.579268
_cons	19.82692	.7326131	27.06	0.000	18.39103	21.26282
var(e.mpg)	27.90954	4.5883			20.22162	38.52027

Linear Regression Path Diagram



Linear Regression Syntax

Syntax using `regress`:

```
regress mpg weight length foreign displacement
```

Syntax using `sem`:

```
sem mpg <- weight length foreign displacement
```

Linear Regression Results

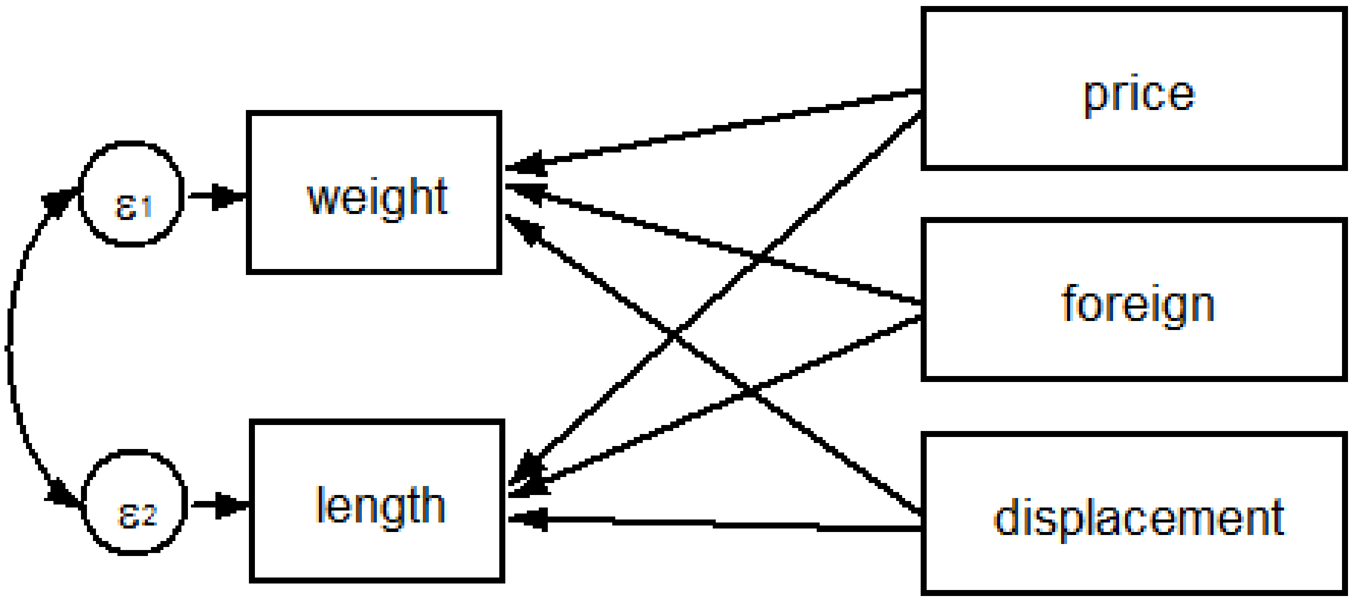
Results using `regress`:

mpg	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
weight	-.0044303	.0019544	-2.27	0.027	-.0083292	-.0005315
length	-.0824511	.0554128	-1.49	0.141	-.1929966	.0280944
foreign	-1.692645	1.105846	-1.53	0.130	-3.898747	.5134562
displacement	.0005878	.0100245	0.06	0.953	-.0194106	.0205861
_cons	50.55702	6.300024	8.02	0.000	37.98882	63.12523

Results using `sem`:

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
mpg <-						
weight	-.0044303	.0018872	-2.35	0.019	-.0081292	-.0007315
length	-.0824511	.053508	-1.54	0.123	-.1873248	.0224226
foreign	-1.692645	1.067833	-1.59	0.113	-3.785559	.400268
displacement	.0005878	.0096799	0.06	0.952	-.0183845	.0195601
_cons	50.55702	6.083464	8.31	0.000	38.63365	62.48039
var(e.mpg)	10.78555	1.773134			7.814581	14.88603

Multivariate Regression Path Diagram



Multivariate Regression Syntax

Syntax using `mvreg`:

```
mvreg weight length = price displacement foreign
```

Syntax using `sem`:

```
sem weight length <- price displacement foreign  ///
    , cov( e.length*e.weight)
```

Multivariate Regression Results

Results using `mvreg`:

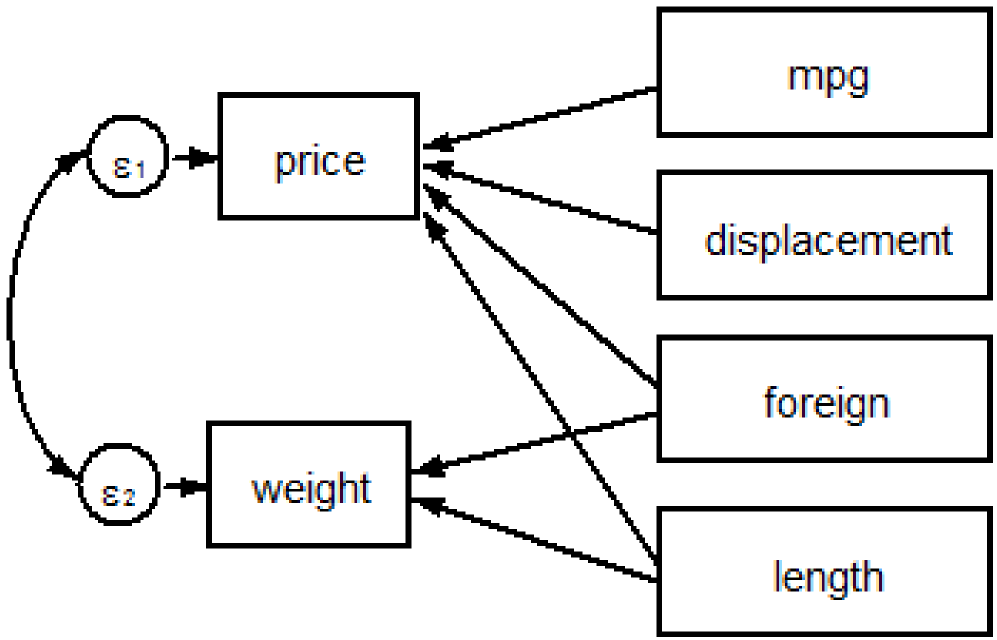
	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
weight						
price	.0570616	.0174226	3.28	0.002	.0223132	.0918099
displacement	5.666956	.7079099	8.01	0.000	4.255074	7.078838
foreign	-324.9114	122.9021	-2.64	0.010	-570.0319	-79.79076
_cons	1646.18	131.626	12.51	0.000	1383.661	1908.7
length						
price	.0006938	.0006547	1.06	0.293	-.0006118	.0019995
displacement	.1699625	.0265999	6.39	0.000	.1169107	.2230143
foreign	-6.988084	4.618077	-1.51	0.135	-16.19855	2.22238
_cons	152.1992	4.945879	30.77	0.000	142.3349	162.0634

Multivariate Regression Results

Results using `sem`:

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
weight <-						
price	.0570616	.0169452	3.37	0.001	.0238496	.0902736
displacement	5.666956	.6885115	8.23	0.000	4.317498	7.016413
foreign	-324.9114	119.5343	-2.72	0.007	-559.1943	-90.6284
_cons	1646.18	128.0192	12.86	0.000	1395.268	1897.093
length <-						
price	.0006938	.0006367	1.09	0.276	-.0005541	.0019418
displacement	.1699625	.025871	6.57	0.000	.1192563	.2206687
foreign	-6.988084	4.49153	-1.56	0.120	-15.79132	1.815153
_cons	152.1992	4.81035	31.64	0.000	142.771	161.6273
var(e.weight)	101330.7	16658.66			73418.28	139854.9
var(e.length)	143.0686	23.52033			103.6591	197.4608
cov(e.weight, e.length)	3133.569	573.2375	5.47	0.000	2010.044	4257.094

Seemingly Unrelated Regression Path Diagram



Seemingly Unrelated Regression Syntax

Syntax using `sureg`:

```
sureg (price foreign mpg displacement)     ///
      (weight foreign length), isure
```

Syntax using `sem`:

```
sem (price <- foreign mpg displacement)   ///
    (weight <- foreign length),           ///
    cov(e.price*e.weight)
```

Seemingly Unrelated Regression Results

Results using `sureg`:

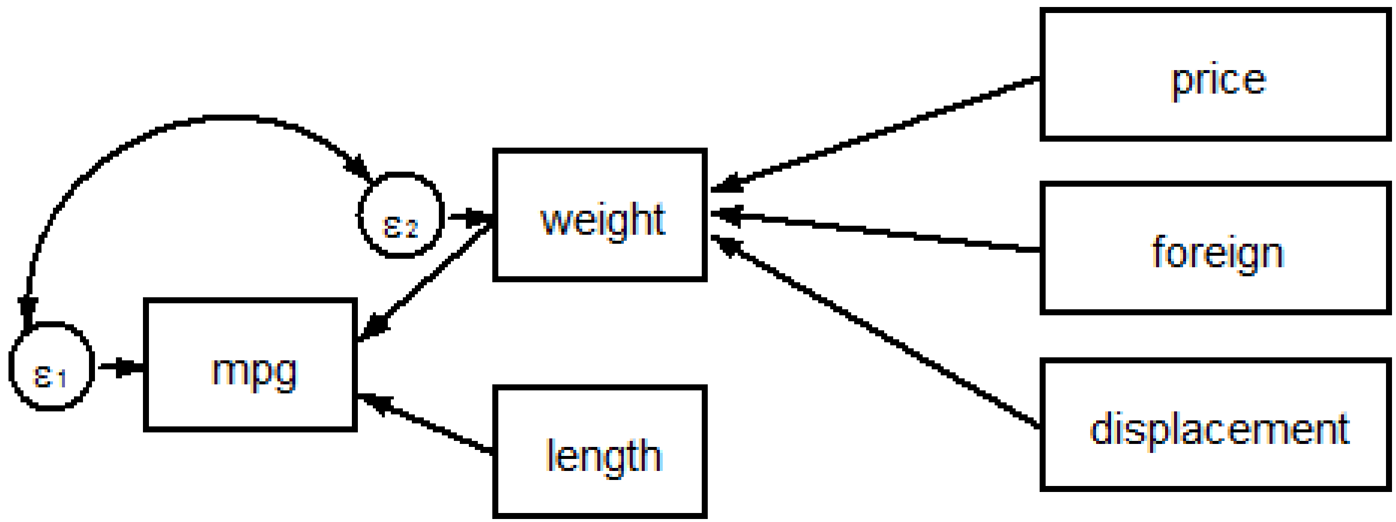
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
price						
foreign	2940.929	691.5458	4.25	0.000	1585.525	4296.334
mpg	-105.0163	57.92716	-1.81	0.070	-218.5514	8.518872
displacement	17.22083	4.244966	4.06	0.000	8.900849	25.54081
_cons	4129.866	1942.567	2.13	0.034	322.5047	7937.228
weight						
foreign	-153.2515	75.33472	-2.03	0.042	-300.9048	-5.598132
length	30.73507	1.528293	20.11	0.000	27.73967	33.73047
_cons	-2711.096	301.6777	-8.99	0.000	-3302.374	-2119.819

Seemingly Unrelated Regression Results

Results using `sem`:

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
price <-						
foreign	2940.929	724.7311	4.06	0.000	1520.482	4361.376
mpg	-105.0163	57.93461	-1.81	0.070	-218.566	8.53347
displacement	17.22083	4.5941	3.75	0.000	8.216558	26.2251
_cons	4129.866	1984.253	2.08	0.037	240.8022	8018.931
weight <-						
foreign	-153.2515	76.21732	-2.01	0.044	-302.6347	-3.868275
length	30.73507	1.584743	19.39	0.000	27.62903	33.84111
_cons	-2711.096	312.6813	-8.67	0.000	-3323.94	-2098.252
var(e.price)	4732491	801783.1			3395302	6596312
var(e.weight)	60253.09	9933.316			43616.45	83235.44
cov(e.price, e.weight)	209268	73909.54	2.83	0.005	64407.92	354128

3-Stage Least Squares Path Diagram



3-Stage Least Squares Syntax

Syntax using `reg3`:

```
reg3 (mpg = weight  length)                    ///
     (weight = price foreign displacement)   ///
      , sure
```

Syntax using `sem`:

```
sem (mpg <- weight length)                         ///
    (weight <- price foreign displacement)         ///
     , cov( e.mpg*e.weight)
```

3-Stage Least Squares Results

Results using `reg3`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mpg						
weight	-.0038705	.0015516	-2.49	0.013	-.0069116	-.0008295
length	-.0752459	.054147	-1.39	0.165	-.181372	.0308802
_cons	47.12534	5.95489	7.91	0.000	35.45397	58.79671
weight						
price	.0566983	.0169217	3.35	0.001	.0235324	.0898642
foreign	-331.9931	119.3554	-2.78	0.005	-565.9254	-98.0608
displacement	5.65145	.6876367	8.22	0.000	4.303707	6.999194
_cons	1653.585	127.892	12.93	0.000	1402.921	1904.248

3-Stage Least Squares Results

Results using `sem`:

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
mpg <-						
weight	-.0038758	.0015516	-2.50	0.012	-.0069168	-.0008347
length	-.0739612	.0550842	-1.34	0.179	-.1819243	.0340019
_cons	46.89969	6.219489	7.54	0.000	34.70971	59.08967
weight <-						
price	.0565862	.0169311	3.34	0.001	.0234018	.0897706
foreign	-334.0496	120.3622	-2.78	0.006	-569.9552	-98.14399
displacement	5.645549	.6888778	8.20	0.000	4.295374	6.995725
_cons	1656.051	129.4058	12.80	0.000	1402.421	1909.682
var(e.mpg)	11.19224	1.842355			8.105893	15.45373
var(e.weight)	101347.1	16664.16			73426.22	139885.1
cov(e.mpg, e.weight)	-76.49352	141.2743	-0.54	0.588	-353.3861	200.3991

Outline

- Introduction to SEM concepts and jargon
- Continuous outcome models using SEM
- **Generalized outcome models using GSEM**
- Multilevel generalized models using GSEM

Generalized outcome models using GSEM

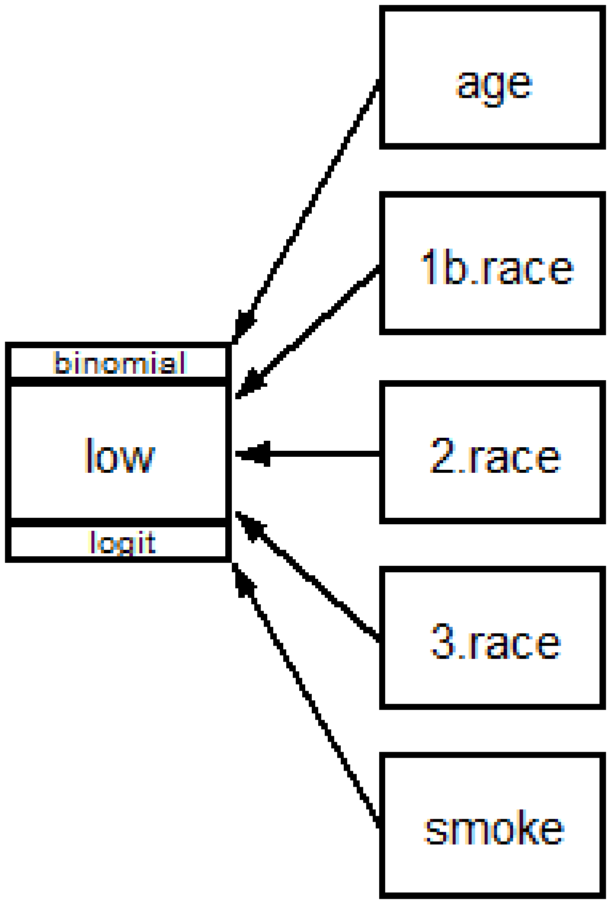
- Logistic regression
- Probit regression
- Multinomial logistic regression
- Ordered logistic regression
- Poisson regression
- Negative binomial regression

Categorical outcome models using GSEM

```
. use "http://www.stata-press.com/data/r13/gsem_lbw", clear
. gen ptl2 = ptl>0
. label var ptl2 "Any history of premature labor"
. recode bwt  (min/2500  = 1 "VeryLow")                 ///
       (2501/3500 = 2 "Low")                            ///
       (3501/max  = 3 "Normal")                         ///
       , gen(bwt_cat)
. label var bwt_cat "Birthweight category"
. describe
```

variable name	storage type	display format	value label	variable label
id	int	%8.0g		subject id
low	byte	%8.0g		birth weight < 2500g
age	byte	%8.0g		age of mother
lwt	int	%8.0g		weight, last menstrual period
race	byte	%8.0g	race	race
smoke	byte	%9.0g	smoke	smoked during pregnancy
ptl	byte	%8.0g		premature labor history (count)
ht	byte	%8.0g		has history of hypertension
ui	byte	%8.0g		presence, uterine irritability
ftv	byte	%8.0g		# physician visits, 1st trimester
bwt	int	%8.0g		birth weight (g)
ptl2	float	%9.0g		Any history of premature labor
bwt_cat	int	%9.0g	bwt_cat	Birthweight category

Logistic Regression Path Diagram



Logistic Regression Syntax

Syntax using `logit` or `logistic`:

```
logit low age i.race smoke

logistic low age i.race smoke
```

Syntax using `gsem`:

```
gsem (low <- age 2.race 3.race smoke,      ///
      family(binomial) link(logit))

gsem low <- age 2.race 3.race smoke, logit

gsem low <- age i.race smoke, logit

estat eform
```

Logistic Regression Results

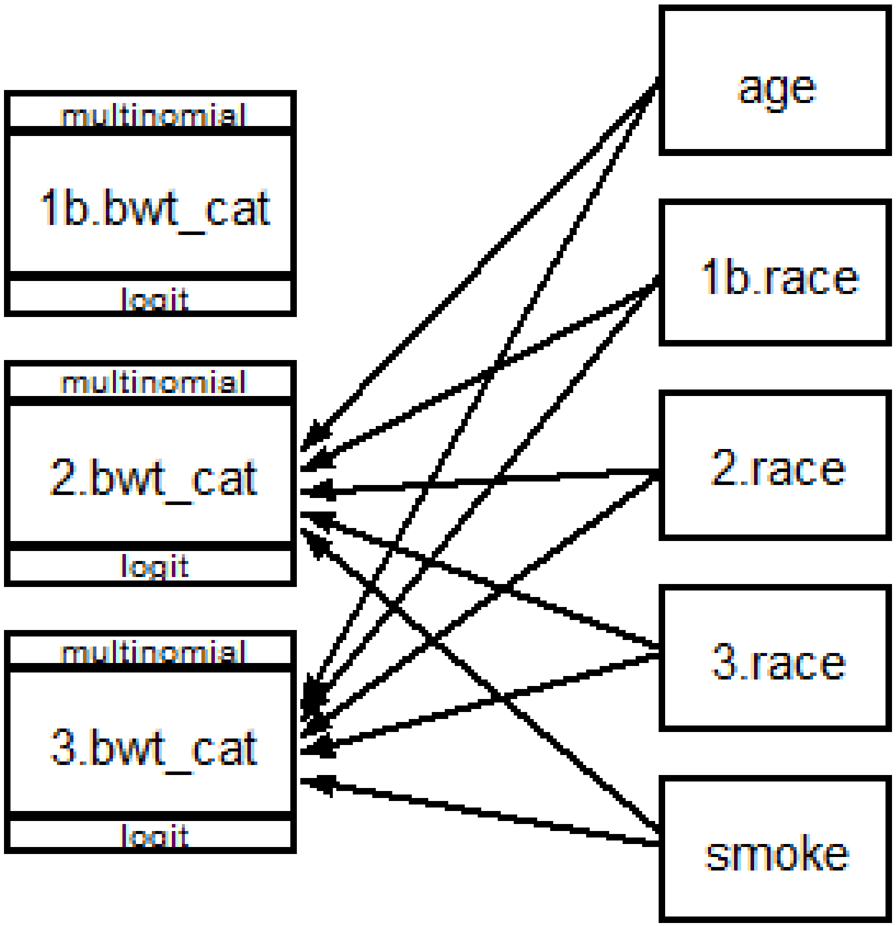
Results using `logistic`:

low	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9657186	.0322573	-1.04	0.296	.9045206	1.031057
race						
black	2.749483	1.356659	2.05	0.040	1.045318	7.231924
other	2.876948	1.167921	2.60	0.009	1.298314	6.375062
smoke	3.00582	1.118001	2.96	0.003	1.449982	6.231081
_cons	.365111	.3146026	-1.17	0.242	.0674491	1.976395

Results using `gsem` and `estat eform`:

low	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9657186	.0322573	-1.04	0.296	.9045206	1.031057
race						
white	1	(empty)				
black	2.749483	1.356659	2.05	0.040	1.045318	7.231924
other	2.876948	1.167921	2.60	0.009	1.298314	6.375062
smoke	3.00582	1.118001	2.96	0.003	1.449982	6.231081
_cons	.365111	.3146026	-1.17	0.242	.0674491	1.976395

Multinomial Logistic Regression Path Diagram



Multinomial Logistic Regression Syntax

Syntax using `mlogit`:

```
mlogit bwt_cat age i.race smoke, baseoutcome(1)
```

Syntax using `gsem`:

```
gsem bwt_cat <- age i.race smoke, mlogit
```

Multinomial Logistic Regression Results

Results using `mlogit`:

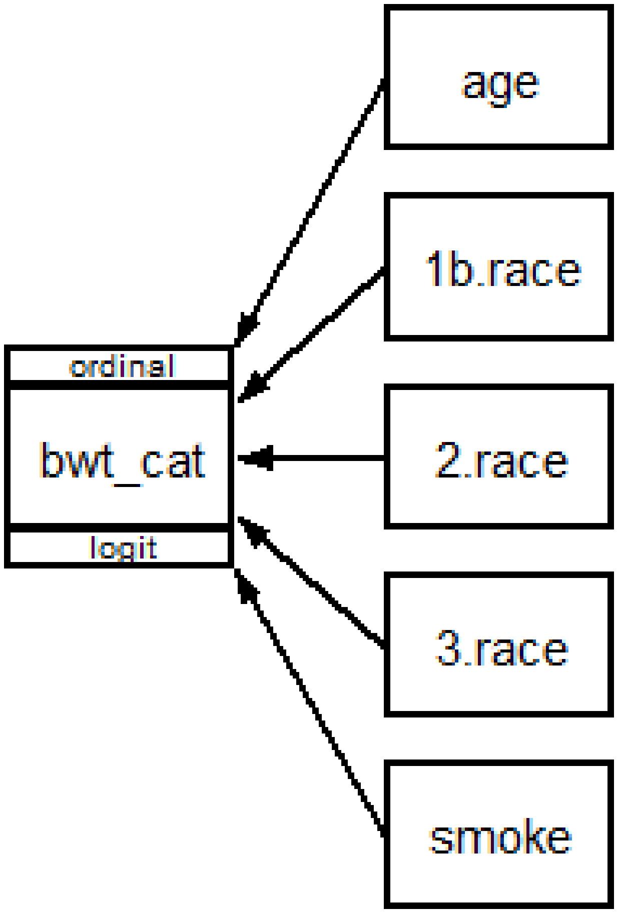
bwt_cat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
VeryLow	(base outcome)					
Low						
age	.0383795	.0349615	1.10	0.272	-.0301437	.1069028
race						
black	-.5139873	.5131883	-1.00	0.317	-1.519818	.4918433
other	-.6468109	.4349866	-1.49	0.137	-1.499369	.2057473
smoke	-.7455204	.3948038	-1.89	0.059	-1.519322	.0282809
_cons	.1573422	.9150125	0.17	0.863	-1.636049	1.950734
Normal						
age	.0133362	.0428736	0.31	0.756	-.0706946	.097367
race						
black	-2.587666	.8727109	-2.97	0.003	-4.298148	-.8771839
other	-2.003564	.5383297	-3.72	0.000	-3.058671	-.948457
smoke	-2.014254	.5166531	-3.90	0.000	-3.026876	-1.001633
_cons	1.155972	1.13431	1.02	0.308	-1.067235	3.379179

Multinomial Logistic Regression Results

Results using `sem`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.bwt_cat	(base outcome)					
2.bwt_cat <-						
age	.0383795	.0349615	1.10	0.272	-.0301437	.1069028
race						
black	-.5139873	.5131883	-1.00	0.317	-1.519818	.4918433
other	-.6468109	.4349866	-1.49	0.137	-1.499369	.2057473
smoke	-.7455204	.3948038	-1.89	0.059	-1.519322	.0282809
_cons	.1573422	.9150125	0.17	0.863	-1.636049	1.950734
3.bwt_cat <-						
age	.0133362	.0428736	0.31	0.756	-.0706946	.097367
race						
black	-2.587666	.8727109	-2.97	0.003	-4.298148	-.8771839
other	-2.003564	.5383297	-3.72	0.000	-3.058671	-.948457
smoke	-2.014254	.5166531	-3.90	0.000	-3.026876	-1.001633
_cons	1.155972	1.13431	1.02	0.308	-1.067235	3.379179

Ordinal Logistic Regression Path Diagram



Ordinal Logistic Regression Syntax

Syntax using `ologit`:

```
ologit bwt_cat age i.race smoke
```

Syntax using `gsem`:

```
gsem bwt_cat <- age 2.race 3.race smoke, ologit
```

Ordinal Logistic Regression Results

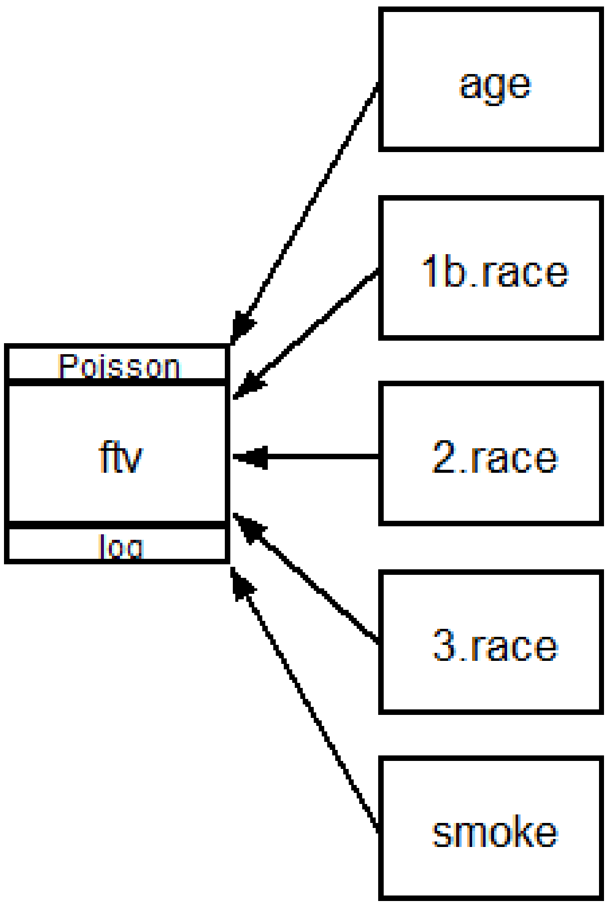
Results using `ologit`:

bwt_cat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0192665	.0268696	0.72	0.473	-.033397	.07193
race						
black	-1.323717	.4380843	-3.02	0.003	-2.182347	-.465088
other	-1.251173	.3399064	-3.68	0.000	-1.917377	-.5849683
smoke	-1.232152	.3152545	-3.91	0.000	-1.850039	-.6142645
/cut1	-1.56245	.7303115			-2.993834	-.1310659
/cut2	.6025144	.7221697			-.8129121	2.017941

Results using `gsem`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
bwt_cat <-						
age	.0192665	.0268696	0.72	0.473	-.033397	.07193
race						
black	-1.323717	.4380843	-3.02	0.003	-2.182347	-.465088
other	-1.251173	.3399064	-3.68	0.000	-1.917377	-.5849683
smoke	-1.232152	.3152545	-3.91	0.000	-1.850039	-.6142645
bwt_cat						
/cut1	-1.56245	.7303115	-2.14	0.032	-2.993834	-.1310659
/cut2	.6025144	.7221697	0.83	0.404	-.8129121	2.017941

Poisson Regression Path Diagram



Poisson Regression Syntax

Syntax using `poisson`:

```
poisson ftv age i.race smoke
```

Syntax using `gsem`:

```
gsem ftv <- age i.race smoke, poisson
```

Poisson Regression Results

Results using `poisson`:

ftv	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0459009	.0144906	3.17	0.002	.0175	.0743019
race						
black	.0336452	.2488814	0.14	0.892	-.4541534	.5214438
other	-.2338308	.1988733	-1.18	0.240	-.6236152	.1559537
smoke	-.1147068	.1775747	-0.65	0.518	-.4627468	.2333332
_cons	-1.216246	.4113805	-2.96	0.003	-2.022537	-.4099547

Results using `gsem`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ftv <-						
age	.0459009	.0144906	3.17	0.002	.0175	.0743019
race						
black	.0336452	.2488814	0.14	0.892	-.4541534	.5214438
other	-.2338308	.1988733	-1.18	0.240	-.6236152	.1559537
smoke	-.1147068	.1775747	-0.65	0.518	-.4627468	.2333332
_cons	-1.216246	.4113805	-2.96	0.003	-2.022537	-.4099547

Outline

- Introduction to SEM concepts and jargon
- Continuous outcome models using SEM
- Generalized outcome models using GSEM
- **Multilevel generalized models using GSEM**

Multilevel generalized outcome models using GSEM

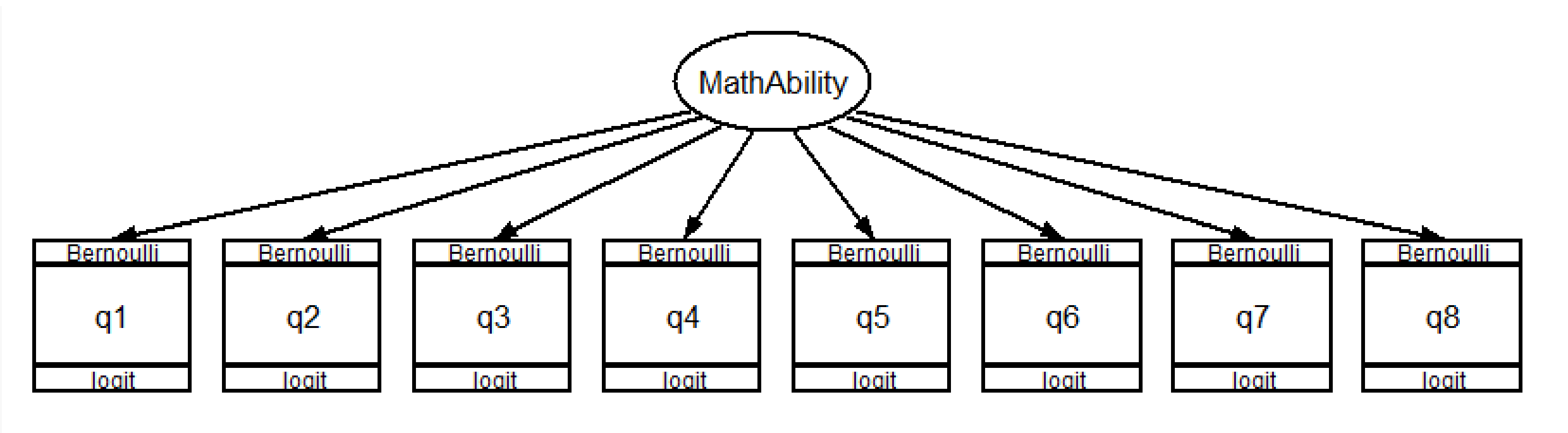
- Measurement component models
- Variance component models
- Latent growth curves
- Latent growth curves for generalized outcomes

Measurement Components Data

```
. use http://www.stata-press.com/data/r13/gsem_cfa, clear
. describe
```

variable name	storage type	display format	value label	variable label
school	byte	%9.0g		School id
id	long	%9.0g		Student id
q1	byte	%9.0g	result	q1 correct
q2	byte	%9.0g	result	q2 correct
q3	byte	%9.0g	result	q3 correct
q4	byte	%9.0g	result	q4 correct
q5	byte	%9.0g	result	q5 correct
q6	byte	%9.0g	result	q6 correct
q7	byte	%9.0g	result	q7 correct
q8	byte	%9.0g	result	q8 correct
att1	float	%26.0g	agree	Skills taught in math class will help me get a better job.
att2	float	%26.0g	agree	Math is important in everyday life
att3	float	%26.0g	agree	Working math problems makes me anxious.
att4	float	%26.0g	agree	Math has always been my worst subject.
att5	float	%26.0g	agree	I am able to learn new math concepts easily.
test1	byte	%9.0g		Score, math test 1
test2	byte	%9.0g		Score, math test 2
test3	byte	%9.0g		Score, math test 3
test4	byte	%9.0g		Score, math test 4

Measurement Component Path Diagram



We can conceptualize the eight measured variables q1-q8 as being realizations of a person's math ability. We can quantify this idea using a latent variable MathAbility.

Measurement Components Syntax

Syntax using `gsem`:

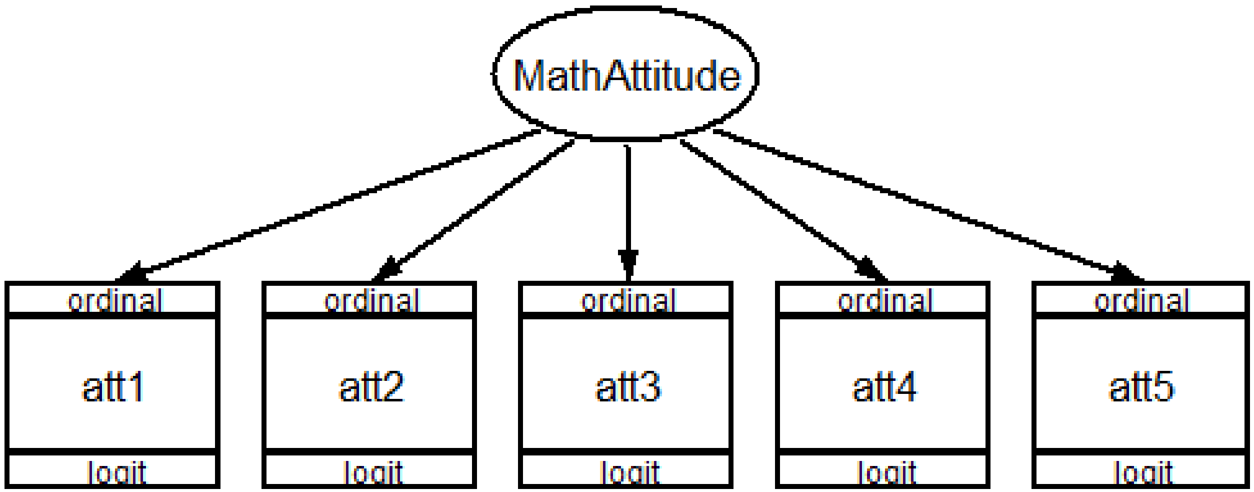
```
gsem (MathAbility -> q1-q8                        ///
      , family(bernoulli) link(logit))            ///
      , latent(MathAbility ) nocapslatent
```

Measurement Components Results

Results using `gsem`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
q1 <-						
MathAbility	1	(constrained)				
_cons	.0373365	.1252279	0.30	0.766	-.2081058	.2827787
q2 <-						
MathAbility	.381626	.116809	3.27	0.001	.1526845	.6105674
_cons	-.4613391	.0989722	-4.66	0.000	-.655321	-.2673571
q3 <-						
MathAbility	.4993762	.134314	3.72	0.000	.2361255	.7626269
_cons	.1533362	.1006072	1.52	0.127	-.0438503	.3505228
q4 <-						
MathAbility	.3299698	.1063034	3.10	0.002	.1216189	.5383207
_cons	-.3230667	.0957983	-3.37	0.001	-.510828	-.1353054
q5 <-						
MathAbility	.8401762	.1995336	4.21	0.000	.4490975	1.231255
_cons	-.0494684	.1163093	-0.43	0.671	-.2774304	.1784937
q6 <-						
MathAbility	.6453722	.1639865	3.94	0.000	.3239646	.9667798
_cons	-.314723	.1083049	-2.91	0.004	-.5269968	-.1024493
q7 <-						
MathAbility	.8163613	.2045448	3.99	0.000	.4154609	1.217262
_cons	.1053404	.1152979	0.91	0.361	-.1206393	.3313201
q8 <-						
MathAbility	.5769516	.1473524	3.92	0.000	.2881463	.865757
_cons	-.026705	.1034396	-0.26	0.796	-.2294429	.1760328
var(MathAbility)	2.151059	.7298407			1.106229	4.182728

Measurement Component Path Diagram



We can also conceptualize the five measured variables att1-att5 as being realizations of a person's attitude about math. We can quantify this idea using a latent variable MathAttitude.

Measurement Components Syntax

Syntax using `gsem`:

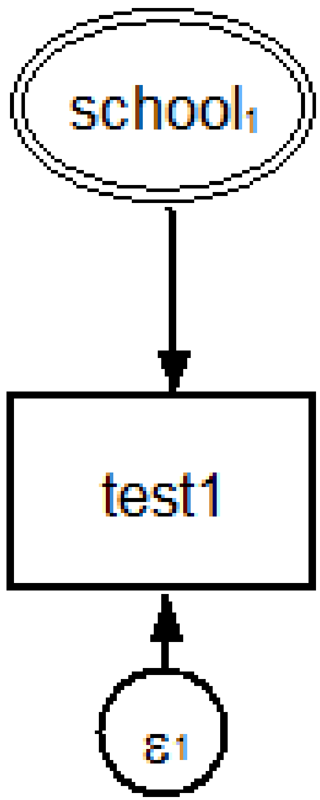
```
gsem (MathAttitude -> att1-att5                    ///
      , family(ordinal) link(logit))               ///
      , latent(MathAttitude ) nocapslatent
```

Measurement Components Results

Results using `gsem`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
att1 <-						
MathAttitude	1	(constrained)				
att2 <-						
MathAttitude	.3651316	.0947737	3.85	0.000	.1793785	.5508846
att3 <-						
MathAttitude	-1.325592	.3281752	-4.04	0.000	-1.968803	-.6823802
att4 <-						
MathAttitude	-.7319336	.1476384	-4.96	0.000	-1.0213	-.4425677
att5 <-						
MathAttitude	.4629576	.1117098	4.14	0.000	.2440104	.6819047
att1						
/cut1	-1.14403	.1407016	-8.13	0.000	-1.4198	-.8682596
/cut2	-.2571716	.1208793	-2.13	0.033	-.4940907	-.0202526
/cut3	.3113316	.121399	2.56	0.010	.0733939	.5492693
/cut4	1.38373	.1505219	9.19	0.000	1.088713	1.678748
att2						
/cut1	-1.058352	.1069425	-9.90	0.000	-1.267955	-.8487485
/cut2	-.1920422	.0946322	-2.03	0.042	-.3775179	-.0065665
/cut3	.3639243	.0957805	3.80	0.000	.1761979	.5516506
/cut4	1.139819	.1090449	10.45	0.000	.9260952	1.353543
att3						
/cut1	-1.003196	.1634751	-6.14	0.000	-1.323601	-.6827905
/cut2	-.0511457	.1372565	-0.37	0.709	-.3201635	.2178721
/cut3	.5278704	.1454233	3.63	0.000	.2428459	.8128949
/cut4	1.587917	.1989801	7.98	0.000	1.197923	1.977911
att4						
/cut1	-1.071316	.1214149	-8.82	0.000	-1.309285	-.8333473
/cut2	-.212007	.1074834	-1.97	0.049	-.4226707	-.0013434
/cut3	.4028505	.1092331	3.69	0.000	.1887576	.6169435
/cut4	1.393148	.1312299	10.62	0.000	1.135942	1.650354
att5						
/cut1	-1.242513	.1147059	-10.83	0.000	-1.467332	-1.017693
/cut2	-.339867	.0983909	-3.45	0.001	-.5327096	-.1470243
/cut3	.2076369	.0974768	2.13	0.033	.0165858	.3986879
/cut4	.9211489	.1067054	8.63	0.000	.7120101	1.130288
var(MathAttit~e)	1.835912	.5279313			1.044917	3.225683

Variance Component Model Path Diagram



Variance Component Model Syntax

Syntax using `mixed`:

```
mixed test1 || school:
```

Syntax using `gsem`:

```
gsem (M1[school] -> test1, ), latent(M1)          ///
      nocapslatent
```

Variance Component Model Results

Results using `mixed`:

test1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	75.548	.6529386	115.70	0.000	74.26826	76.82774

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
school: Identity				
var(_cons)	7.410363	2.697303	3.630865	15.12408
var(Residual)	27.90533	1.801282	24.58909	31.66883

Results using `gsem`:

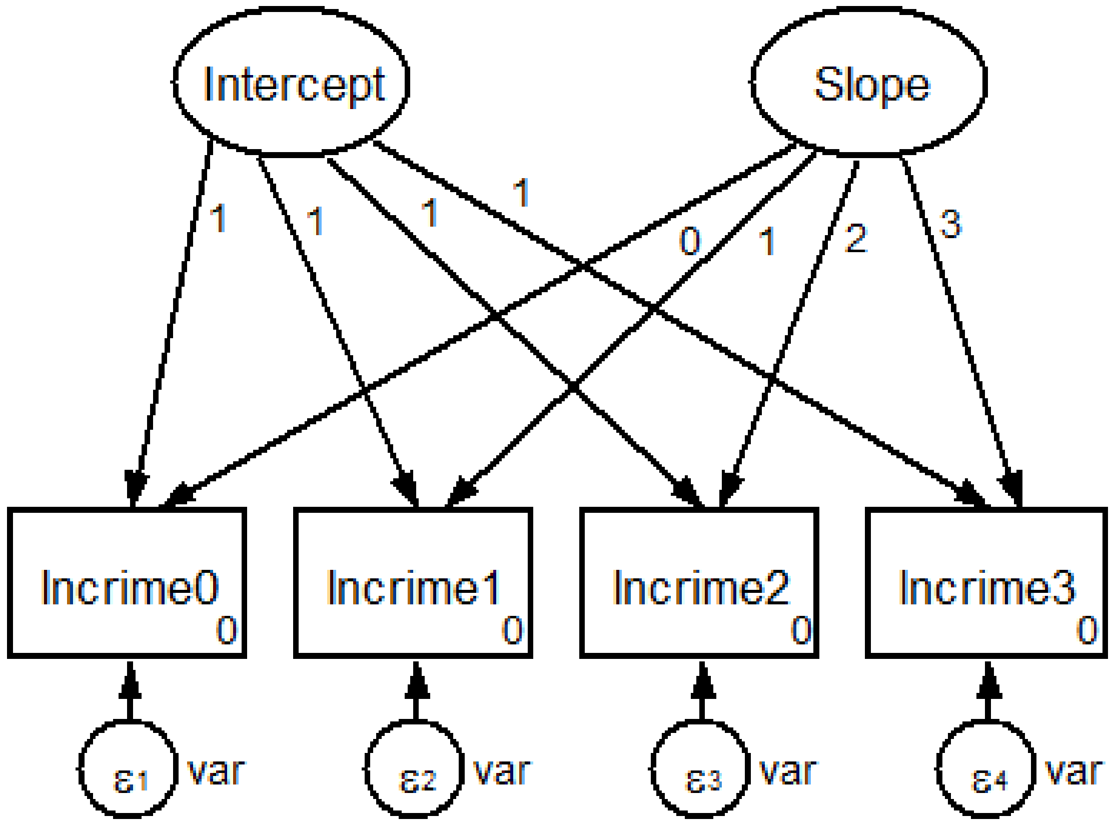
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
test1 <-						
M1[school]	1	(constrained)				
_cons	75.548	.6529386	115.70	0.000	74.26826	76.82774
var(M1[school])	7.410362	2.697302			3.630865	15.12407
var(e.test1)	27.90533	1.801281			24.58908	31.66882

Latent Growth Curve Data

```
. use http://www.stata-press.com/data/r13/sem_lcm, clear
. gen id = _n
. describe
```

variable name	storage type	display format	value label	variable label
lncrime0	float	%9.0g		ln(crime rate) in Jan & Feb
lncrime1	float	%9.0g		ln(crime rate) in Mar & Apr
lncrime2	float	%9.0g		ln(crime rate) in May & Jun
lncrime3	float	%9.0g		ln(crime rate) in Jul & Aug
id	float	%9.0g		

Latent Growth Curve Path Diagram



Latent Growth Curve Syntax

Syntax using `mixed`:

```
gen id = _n
reshape long lncrime, i(id) j(time)
mixed lncrime time || id: time, cov(unstr)
```

Syntax using `sem`:

```
reshape wide
sem (lncrime0 <- Intercept@1 Slope@0 _cons@0)  ///
    (lncrime1 <- Intercept@1 Slope@1 _cons@0)  ///
    (lncrime2 <- Intercept@1 Slope@2 _cons@0)  ///
    (lncrime3 <- Intercept@1 Slope@3 _cons@0), ///
    latent(Intercept Slope)                    ///
    var(e.lncrime0@var e.lncrime1@var          ///
      e.lncrime2@var e.lncrime3@var)           ///
    means(Intercept Slope)                     ///
    nocnsreport nolog
```

Latent Growth Curve Results

Results using `mixed`:

lncrime	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	.1426952	.0104574	13.65	0.000	.1221992	.1631912
_cons	5.337915	.0407501	130.99	0.000	5.258047	5.417784

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
id: Unstructured				
var(time)	.0196198	.0031082	.0143829	.0267635
var(_cons)	.5274091	.0446436	.4467824	.6225859
cov(time,_cons)	-.034316	.0088848	-.0517298	-.0169022
var(Residual)	.0981956	.0051826	.0885456	.1088972

Results using `sem`:

mean(Intercept)	5.337915	.0407501	130.99	0.000	5.258047	5.417784
mean(Slope)	.1426952	.0104574	13.65	0.000	.1221992	.1631912
var(e.lncrime0)	.0981956	.0051826			.0885457	.1088972
var(e.lncrime1)	.0981956	.0051826			.0885457	.1088972
var(e.lncrime2)	.0981956	.0051826			.0885457	.1088972
var(e.lncrime3)	.0981956	.0051826			.0885457	.1088972
var(Intercept)	.527409	.0446436			.4467822	.6225858
var(Slope)	.0196198	.0031082			.0143829	.0267635
cov(Intercept, Slope)	-.034316	.0088848	-3.86	0.000	-.0517298	-.0169022

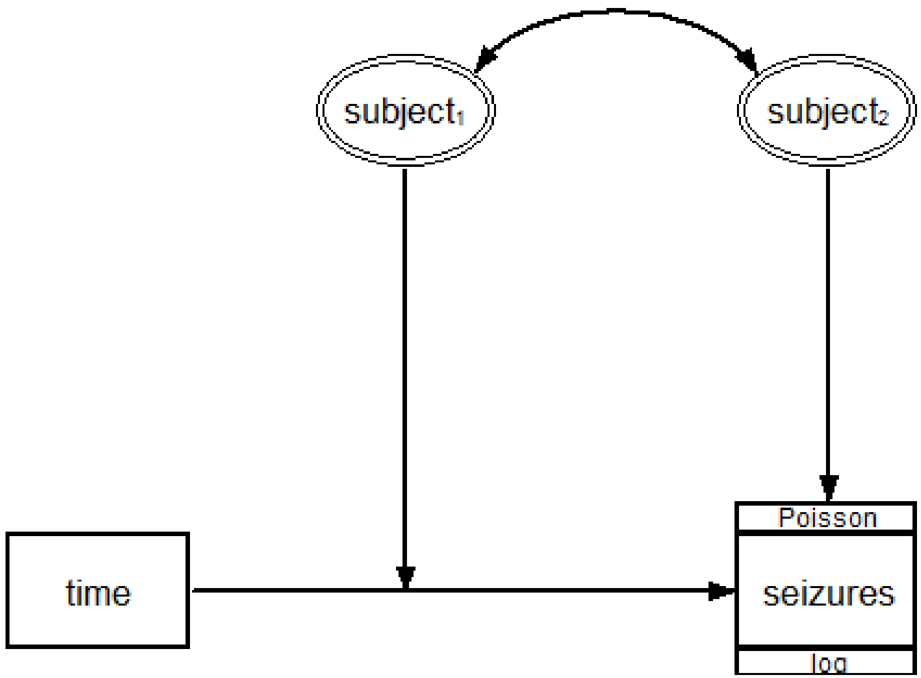
Poisson Latent Growth Curve Data

(data in long format)

variable name	storage type	display format	value label	variable label
subject	byte	%9.0g		Subject ID: 1-59
seizures	int	%9.0g		No. of seizures
treat	byte	%9.0g		1: progabide; 0: placebo
visit	float	%9.0g		Dr. visit; coded as (-.3, -.1, .1, .3)
lage	float	%9.0g		log(age), mean-centered
lbas	float	%9.0g		log(0.25*baseline seizures), mean-centered
lbas_trt	float	%9.0g		lbas/treat interaction
v4	byte	%8.0g		Fourth visit indicator
time	float	%9.0g		

Sorted by: subject visit

Poisson Latent Growth Curve Path Diagram
(data in long format)



Poisson Latent Growth Curve Path Diagram
(data in long format)

Syntax using `mepoisson`:

```
mepoisson seizures time || subject: time, cov(unstr)
```

Syntax using `gsem`:

```
gsem (seizures <- time c.time#S[subject] I[subject]), ///
     family(poisson) link(log)
```

Poisson Latent Growth Curve Path Diagram
(data in long format)

Results using `mepoisson`:

seizures	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time	-.0503416	.0355162	-1.42	0.156	-.119952	.0192689
_cons	1.682242	.1386123	12.14	0.000	1.410566	1.953917
subject						
var(time)	.0211453	.0091765			.0090326	.0495009
var(_cons)	.9545034	.2064406			.6247109	1.458397
subject						
cov(_cons,time)	-.0362171	.0337022	-1.07	0.283	-.1022722	.0298379

Poisson Latent Growth Curve Path Diagram
(data in long format)

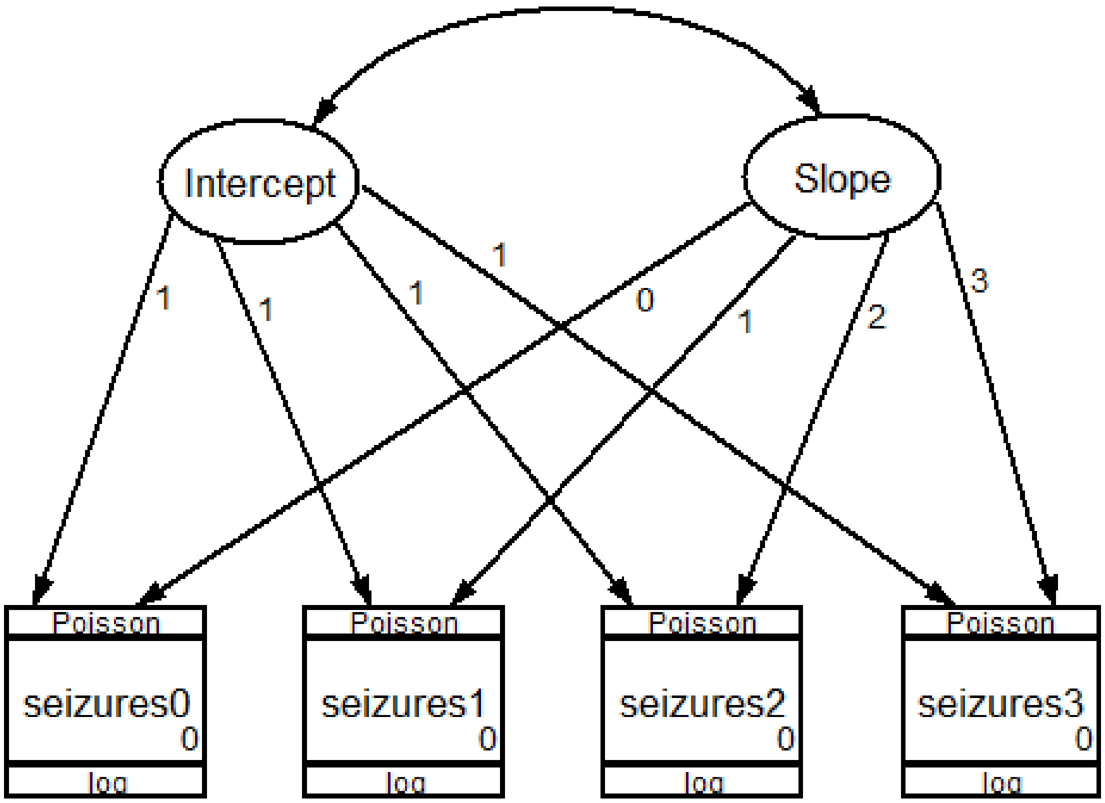
Results using `gsem`:

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
seizures <-						
time	-.0503416	.0355162	-1.42	0.156	-.119952	.0192689
c.time# S[subject]	1	(constrained)				
I[subject]	1	(constrained)				
_cons	1.682242	.1386123	12.14	0.000	1.410566	1.953917
var(S[subject])	.0211453	.0091765			.0090326	.0495009
var(I[subject])	.9545034	.2064406			.6247109	1.458397
cov(I[subject], S[subject])	-.0362171	.0337022	-1.07	0.283	-.1022722	.0298379

Poisson Latent Growth Curve Data
(data in wide format)

```
--------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
subject         byte    %9.0g                 Subject ID: 1-59
seizures0       int     %9.0g                 0 seizures
lage            float   %9.0g                 0 lage
seizures1       int     %9.0g                 1 seizures
seizures2       int     %9.0g                 2 seizures
seizures3       int     %9.0g                 3 seizures
--------------------------------------------------------------------------------
Sorted by:  subject
```

Poisson Latent Growth Curve Path Diagram
(data in wide format)



Poisson Latent Growth Curve Path Diagram
(data in wide format)

Syntax using `gsem`:

```
gsem (seizures0 <- Intercept@1 Slope@0)          ///
     (seizures1 <- Intercept@1 Slope@1)          ///
     (seizures2 <- Intercept@1 Slope@2)          ///
     (seizures3 <- Intercept@1 Slope@3),         ///
            nocons means(Intercept Slope)        ///
            family(poisson) link(log)
```

Poisson Latent Growth Curve Path Diagram
(data in wide format)

Results using `gsem`:

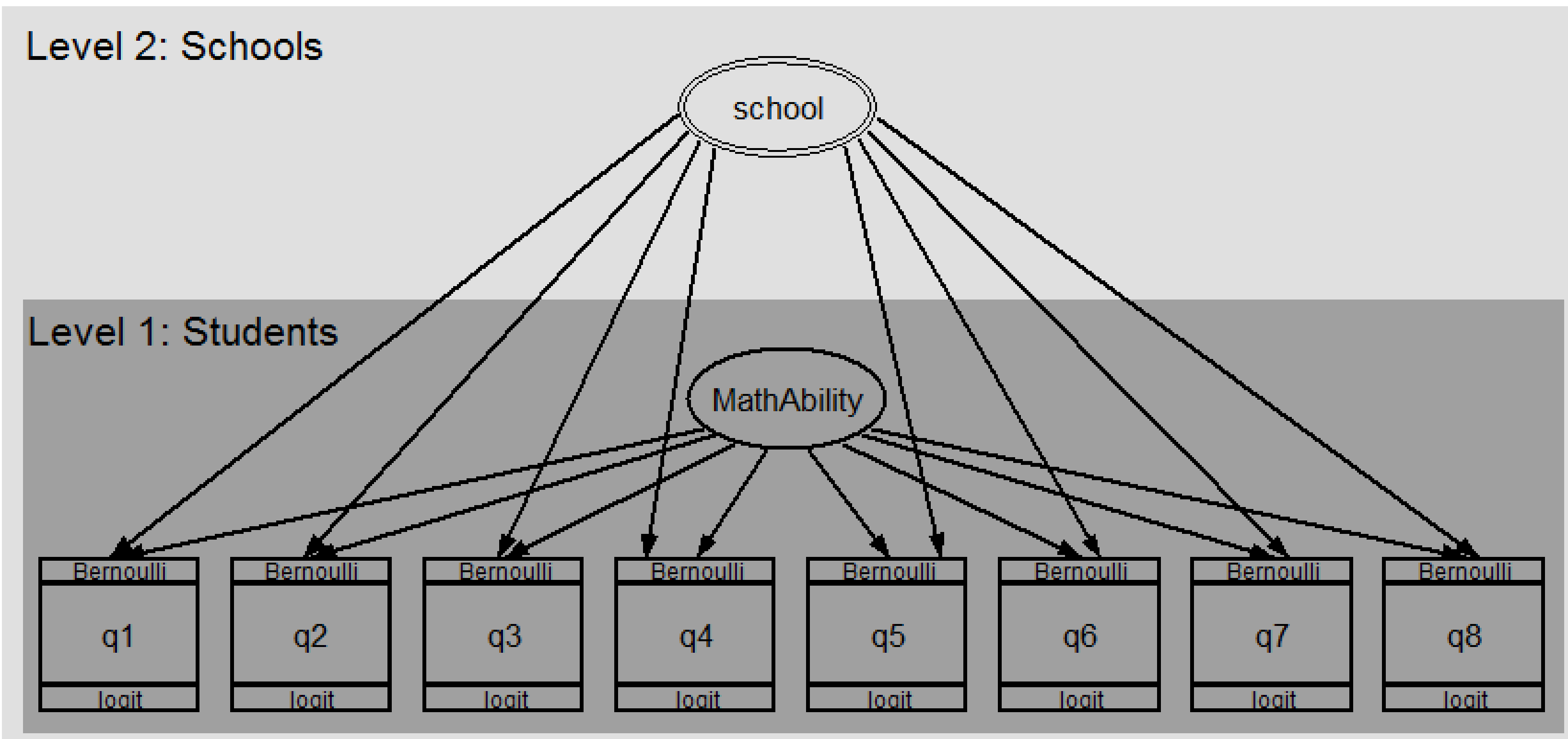
	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
seizures0 <-						
Intercept	1	(constrained)				
Slope	0	(omitted)				
_cons	0	(omitted)				
seizures1 <-						
Intercept	1	(constrained)				
Slope	1	(constrained)				
_cons	0	(omitted)				
seizures2 <-						
Intercept	1	(constrained)				
Slope	2	(constrained)				
_cons	0	(omitted)				
seizures3 <-						
Intercept	1	(constrained)				
Slope	3	(constrained)				
_cons	0	(omitted)				
mean(Intercept)	1.682215	.138856	12.11	0.000	1.410062	1.954368
mean(Slope)	-.0503355	.0355747	-1.41	0.157	-.1200606	.0193896
var(Intercept)	.9543458	.2067282			.6241952	1.45912
var(Slope)	.0211439	.0091791			.0090294	.0495122
cov(Slope, Intercept)	-.0362289	.0339178	-1.07	0.285	-.1027065	.0302486

Multilevel GSEM Data

```
. use http://www.stata-press.com/data/r13/gsem_cfa, clear
. describe
```

variable name	storage type	display format	value label	variable label
school	byte	%9.0g		School id
id	long	%9.0g		Student id
q1	byte	%9.0g	result	q1 correct
q2	byte	%9.0g	result	q2 correct
q3	byte	%9.0g	result	q3 correct
q4	byte	%9.0g	result	q4 correct
q5	byte	%9.0g	result	q5 correct
q6	byte	%9.0g	result	q6 correct
q7	byte	%9.0g	result	q7 correct
q8	byte	%9.0g	result	q8 correct
att1	float	%26.0g	agree	Skills taught in math class will help me get a better job.
att2	float	%26.0g	agree	Math is important in everyday life
att3	float	%26.0g	agree	Working math problems makes me anxious.
att4	float	%26.0g	agree	Math has always been my worst subject.
att5	float	%26.0g	agree	I am able to learn new math concepts easily.
test1	byte	%9.0g		Score, math test 1
test2	byte	%9.0g		Score, math test 2
test3	byte	%9.0g		Score, math test 3
test4	byte	%9.0g		Score, math test 4

Multilevel GSEM Path Diagram



Multilevel GSEM Syntax

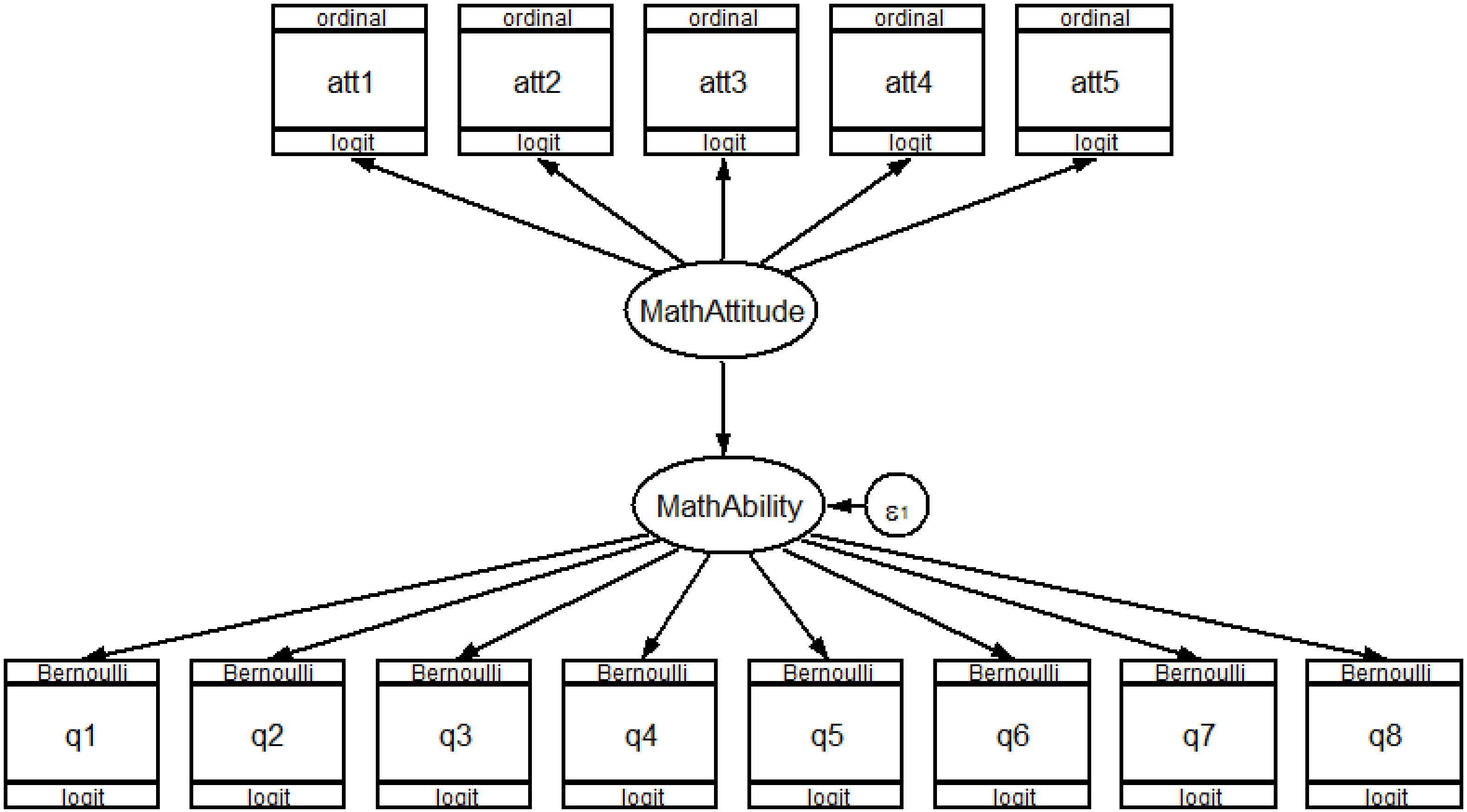
Syntax using `gsem`:

```
gsem (MathAbility -> q1-q8, family(bernoulli) link(logit))     ///
     (School[school] -> q1-q8, family(bernoulli) link(logit))  ///
      , latent(MathAbility School)                              ///
      covstruct(_lexogenous, diagonal)                          ///
      nocapslatent
```

Generalized Structural Equation Models

We can combine measurement components to fit a dizzying variety of models that can simultaneously combine longitudinal, latent growth curve and multilevel structures that cannot be modeled with other Stata commands.

Multilevel GSEM Path Diagram

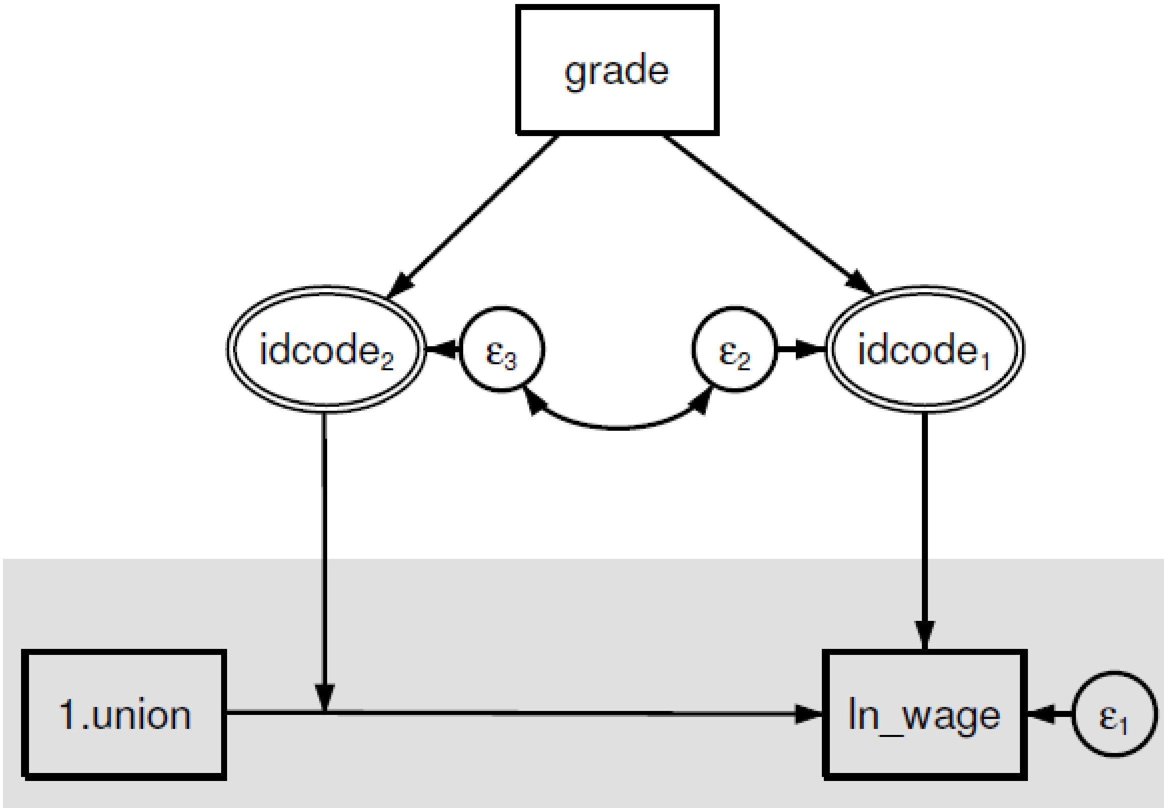


Multilevel GSEM Syntax

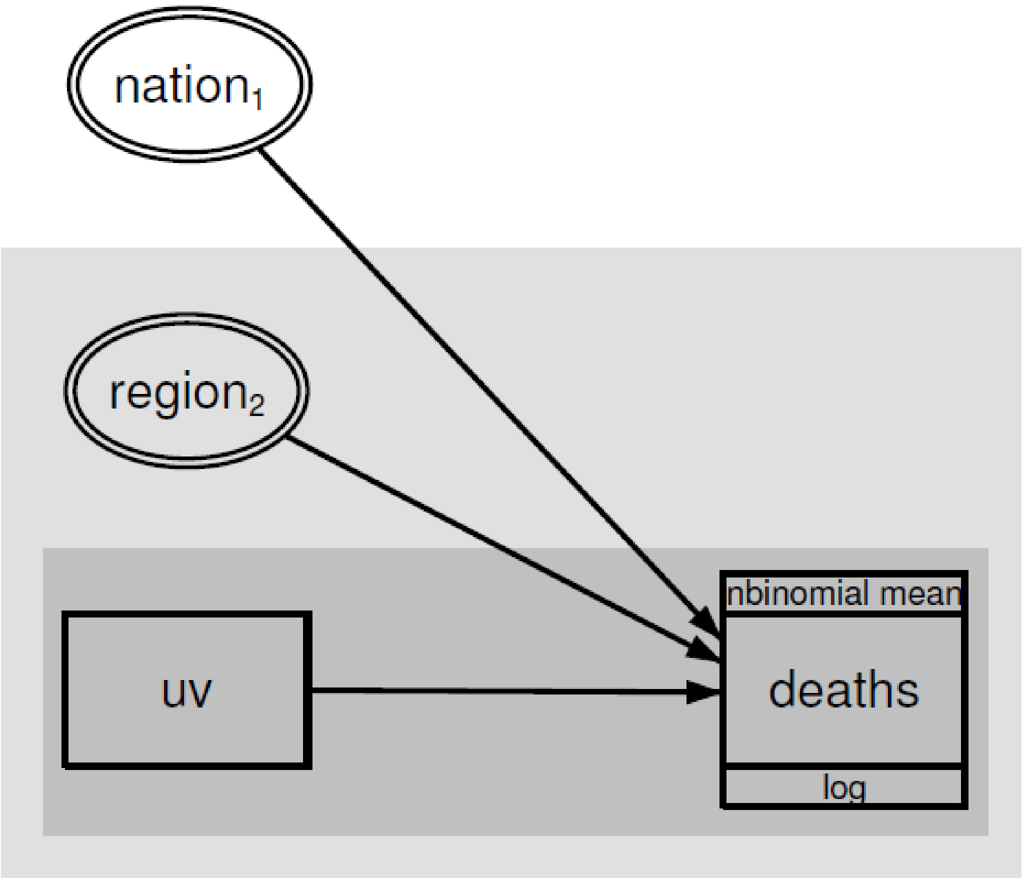
Syntax using `gsem`:

```
gsem (MathAbility -> q1-q8, family(bernoulli) link(logit))      ///
     (MathAttitude -> MathAbility, )                            ///
     (MathAttitude -> att1-att5, family(ordinal) link(logit))   ///
      , latent(MathAbility MathAttitude ) nocapslatent
```

GSEM Examples

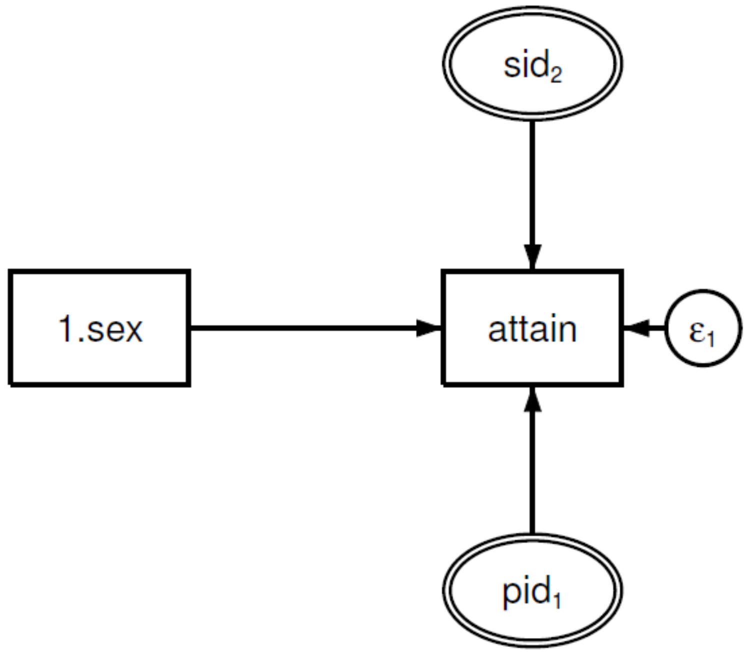


GSEM Examples



Three level GSEM

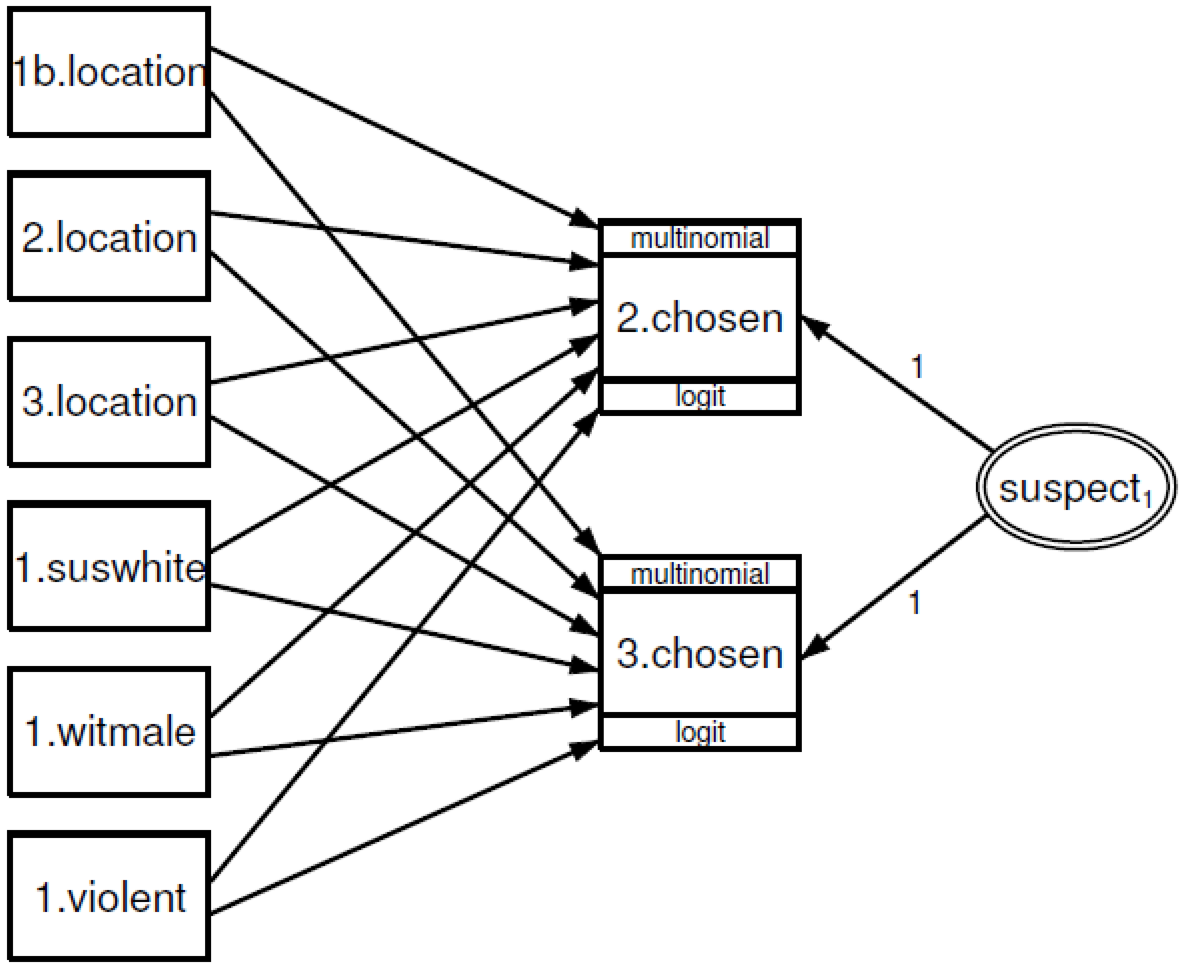
GSEM Examples



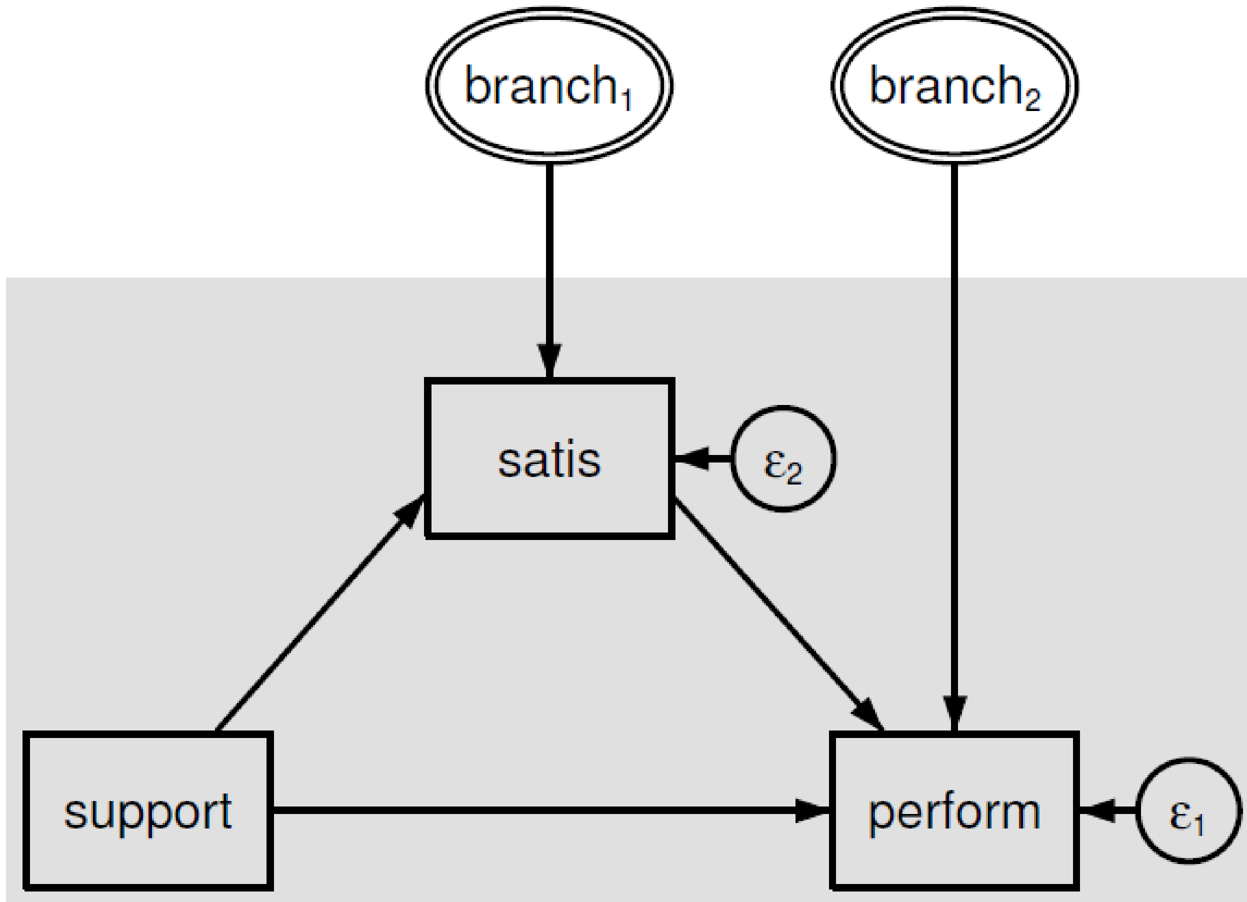
Crossed Model: pid1 is the students' primary school ID and sid2 is the students' secondary school ID number. The two multilevel latent variables account for the nesting of students within each of the two schools.

Data Analysis and Statistical Software

GSEM Examples



GSEM Examples



Multilevel Mediation Model

Conclusions

- Many regression models are a special case of SEM/GSEM
- SEM/GSEM allow us to fit complex structural equation models that we cannot fit with other regression techniques
- Stata's SEM Builder makes it easy to draw and estimate structural equation models

References and Further Reading

Stata 13 Structural Equation Modeling Reference Manual:
www.stata.com/manuals13/sem.pdf

Acock, A.C. (2013) Discovering Structural Equation Modeling Using Stata, Revised Edition . College Station, TX: Stata Press.

Rabe-Hesketh, S., and A. Skrondal. (2012) Multilevel and Longitudinal Modeling Using Stata. 3rd ed. College Station, TX: Stata Press.