Estimating and interpreting structural equation models in Stata 12

David M. Drukker

Director of Econometrics
Stata

Stata Conference, Chicago
July 14, 2011

Purpose and outline

Purpose
To excite structural-equation-model (SEM) devotees by describing part of the new `sem` command and convince traditional simultaneous-equation-model types that the `sem` command is worth investigating

Outline

1. The language of SEM

2. Parameter estimation
   - SUR with observed exogenous variables
   - Recursive (triangular) system with correlated errors
   - SUR with observed exogenous variables and a latent variable
   - Nonrecursive system with a latent variable

3. Postestimation

Variables and Paths

There are five types of variables in SEMs

- A variable is either observed or latent
  - Observed variables are in your dataset
  - Unobserved variables are not in your dataset, but you wish they were

- A variable is either exogenous or endogenous
  - A variable is exogenous if it is determined outside the system
  - A variable is endogenous if is not exogenous

- The concepts give rise to four possibilities
  - Observed exogenous variable, latent exogenous variable, observed endogenous variable, and latent endogenous variable

- Errors are a special type of latent exogenous variables
  - Errors are the random shocks or effects that drive the system
  - Errors are the random effects that cause the outcomes of "observationally equivalent individuals" to differ

Path diagram

- A path diagram is graphical specification of model
- A path diagram is composed of
  - Variables in square or rectangular boxes are observed variables
  - Variables in circles or ellipses are latent variables
  - Straight arrows
    - Each straight arrow indicates that the variable at the base affects the variable at the head
    - When two variables have two arrows that point to each other there is feedback; each one affects the other
  - Curved two-headed arrows indicate that two variables are correlated
  - A number along an arrow represents a constraint

Path diagram
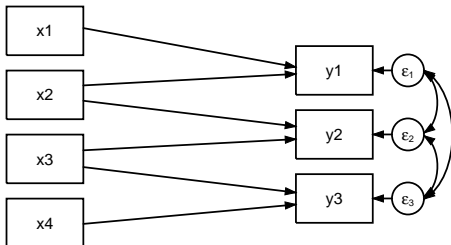
- This is a path diagram for a seemingly unrelated regression (SUR) model with observed exogenous variables

Mathematical description of model

- SUR with observed exogenous variables

$$y_1 = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \epsilon_1$$
$$y_2 = \beta_{20} + \beta_{22}x_2 + \beta_{23}x_3 + \epsilon_2$$
$$y_3 = \beta_{30} + \beta_{33}x_2 + \beta_{34}x_4 + \epsilon_3$$

where $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)'$, $\mathbf{E}[\epsilon] = (0, 0, 0)'$, and $\mathbf{Var}[\epsilon] = \mathbf{\Sigma}$

- ```
  sem (y1 <- x1 x2) (y2 <- x2 x3) (y3 <- x3 x4) ,
  cov(e.y2*e.y1 e.y3*e.y2 e.y3*e.y1)
  ```
  alternatively
  ```
  sem (y1 <- x1 x2) (y2 <- x2 x3) (y3 <- x3 x4) ,
  covstructure(e._Endogenous, unstructured)
  ```

Estimate SUR by sem

```
. sem   (y1 <- x1 x2) (y2 <- x2 x3) (y3 <- x3 x4) ,        ///
>       covariance(e.y2*e.y1 e.y3*e.y2 e.y3*e.y1) nolog
Endogenous variables
Observed:  y1 y2 y3
Exogenous variables
Observed:  x1 x2 x3 x4
Structural equation model                       Number of obs      =       500
Estimation method = ml
Log likelihood    = -6783.5255
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
Structural						
y1 <-						
x1	.9856651	.0349005	28.24	0.000	.9172614	1.054069
x2	.5498082	.0411897	13.35	0.000	.4690778	.6305385
_cons	.9780043	.0827437	11.82	0.000	.8158297	1.140179
y2 <-						
x2	.3666458	.0443247	8.27	0.000	.2797711	.4535206
x3	1.088846	.0402088	27.08	0.000	1.010038	1.167654
_cons	-1.002962	.0843895	-11.88	0.000	-1.168363	-.837562
y3 <-						
x3	.3069075	.0408562	7.51	0.000	.2268308	.3869841
x4	.7640136	.0396892	19.25	0.000	.6862241	.841803
_cons	1.044546	.0874646	11.94	0.000	.8731183	1.215973
Variance						
e.y1	3.408108	.2158503			3.010253	3.858545
e.y2	3.545391	.2244101			3.131744	4.013674
e.y3	3.823403	.242093			3.377172	4.328596
Covariance						
e.y1						
e.y2	1.949872	.1785632	10.92	0.000	1.599895	2.29985
e.y3	2.151246	.1884359	11.42	0.000	1.781918	2.520573
e.y2						
e.y3	1.940438	.1866187	10.40	0.000	1.574672	2.306204

```
LR test of model vs. saturated: chi2(6)   =      7.40, Prob > chi2 = 0.2855
. estimates store sur_sem
```

Estimate SUR by sureg

```
. sureg (y1 = x1 x2) (y2 = x2 x3) (y3 = x3 x4) , isure nolog  tol(1e-15)
Seemingly unrelated regression, iterated
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
y1	500	2	1.846106	0.6512	1447.37	0.0000
y2	500	2	1.882921	0.6335	1169.28	0.0000
y3	500	2	1.955352	0.4582	644.10	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
y1						
x1	.9856651	.0348271	28.30	0.000	.9174052	1.053925
x2	.5498082	.0411671	13.36	0.000	.4691222	.6304941
_cons	.9780043	.0827435	11.82	0.000	.81583	1.140179
y2						
x2	.3666458	.0442686	8.28	0.000	.279881	.4534107
x3	1.088846	.0401428	27.12	0.000	1.010167	1.167524
_cons	-1.002962	.084389	-11.88	0.000	-1.168362	-.8375629
y3						
x3	.3069075	.0407619	7.53	0.000	.2270156	.3867993
x4	.7640136	.0395484	19.32	0.000	.6865001	.841527
_cons	1.044546	.0874645	11.94	0.000	.8731185	1.215973

```
. estimates store sur_sureg
```

Results are the same

```
. estimates table sur_sem sur_sureg, b se(%7.6g) keep(y1: y2: y3:)
```

Variable	sur_sem	sur_sureg
y1		
x1	.98566508	.98566508
	.0349	.03483
x2	.54980818	.54980818
	.04119	.04117
_cons	.97800427	.97800427
	.08274	.08274
y2		
x2	.36664584	.36664584
	.04432	.04427
x3	1.0888457	1.0888457
	.04021	.04014
_cons	-1.0029623	-1.0029623
	.08439	.08439
y3		
x3	.30690746	.30690746
	.04086	.04076
x4	.76401355	.76401355
	.03969	.03955
_cons	1.0445458	1.0445458
	.08746	.08746

legend: b/se

Sembuilder

- There is an awesome GUI for **sem**
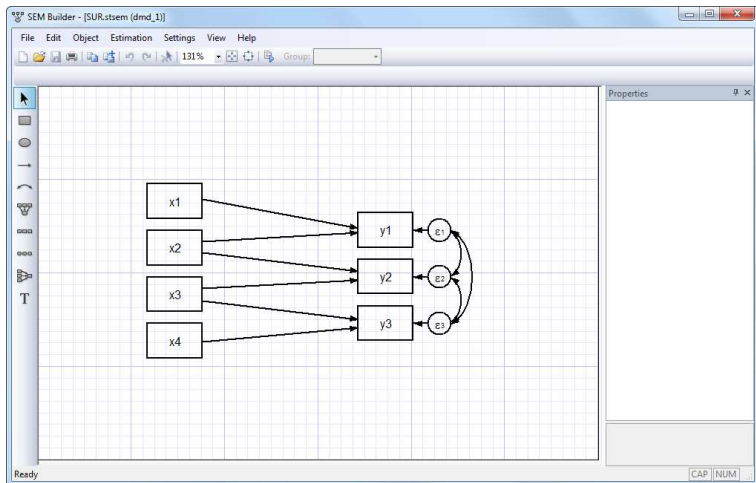
Covariates, errors, and distributions

In all the examples that I discuss

- The analysis is conditional on the exogenous variables
- We assume that the vector of errors, denoted by ϵ, is independently and identically distributed over the observations
- We do not need to assume that the ϵ is normally, or even symmetrically distributed
- Both the Maximum Likelihood (ML) and the asymptotically distribution free (ADF) estimators are consistent and asymptotically normally distributed

    - Specify vce(robust) with the ML estimator, if the ϵ are not assumed to be normally distributed
    - If the ϵ are normally distributed, the ML estimator is more efficient than the ADF estimator
    - The ADF estimator is a generalized method of moments (GMM) estimator

Path diagram
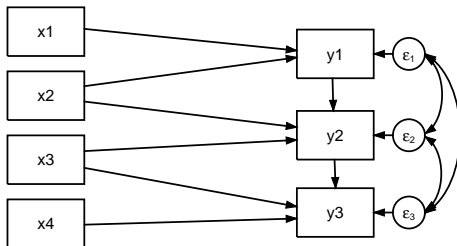
- Recursive system with correlated errors (SEM language)
    - Sometimes called partially recursive system with correlated errors (SEM language)
- Triangular system with correlated errors (Econometric language)
- The system of equations has a recursive structure, but the errors are correlated so the equation-by-equation ordinary least-squares (OLS) estimator is not consistent.

Mathematical description of model

- Recursive (triangular) system with correlated errors

$$y_1 = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \epsilon_1$$
$$y_2 = \beta_{20} + \gamma_{21}y_1 + \beta_{22}x_2 + \beta_{23}x_3 + \epsilon_2$$
$$y_3 = \beta_{30} + \gamma_{32}y_2 + \beta_{33}x_2 + \beta_{34}x_4 + \epsilon_3$$

where $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)'$, $\mathbf{E}[\epsilon] = (0, 0, 0)'$, and $\mathbf{Var}[\epsilon] = \mathbf{\Sigma}$

- ```
  sem (y1 <- x1 x2) (y2 <- y1 x2 x3) (y3 <- y2 x3 x4) ,
  cov(e.y2*e.y1 e.y3*e.y2 e.y3*e.y1)
  ```

Estimate recursive model by sem

```
. sem   (y1 <- x1 x2) (y2 <- y1 x2 x3) (y3 <- y2 x3 x4) ,          ///
>       cov(e.y2*e.y1 e.y3*e.y2 e.y3*e.y1) nolog
Endogenous variables
Observed:  y1 y2 y3
Exogenous variables
Observed:  x1 x2 x3 x4
Structural equation model                      Number of obs     =        500
Estimation method = ml
Log likelihood    = -6882.468
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
Structural						
y2 <-						
y1	.5160285	.0463833	11.13	0.000	.4251189	.6069381
x2	.5097059	.0627235	8.13	0.000	.38677	.6326417
x3	1.009926	.0429949	23.49	0.000	.9256576	1.094194
_cons	-1.027349	.0983129	-10.45	0.000	-1.220039	-.8346598
y3 <-						
y2	.5732565	.0454244	12.62	0.000	.4842263	.6622868
x3	.2917948	.0729249	4.00	0.000	.1488646	.434725
x4	.8197978	.0444761	18.43	0.000	.7326262	.9069694
_cons	.8690175	.0896196	9.70	0.000	.6933663	1.044669
y1 <-						
x1	.992947	.0388631	25.55	0.000	.9167763	1.069118
x2	.5402264	.0417589	12.94	0.000	.4583805	.6220723
_cons	.8546342	.0775166	11.03	0.000	.7027045	1.006564
Variance						
e.y1	2.988624	.1890178			2.640197	3.383033
e.y2	3.886285	.2900963			3.357343	4.498561
e.y3	3.563744	.3279422			2.97562	4.26811
Covariance						
e.y1						
e.y2	1.669049	.2188169	7.63	0.000	1.240176	2.097922
e.y3	1.592503	.2179365	7.31	0.000	1.165355	2.019651
e.y2						
e.y3	1.805499	.3037502	5.94	0.000	1.21016	2.400839

```
LR test of model vs. saturated: chi2(4)   =      0.04, Prob > chi2 = 0.9998
. estimates store sur_sem
```

Estimate recursive model by GLS

- There is a long history in statistics and econometrics of "tricking" readily available estimators to handle more complicated problems
- Using a generalized least squares (GLS) estimator of a triangular SUR model to estimate the parameters of triangular models goes back to [Lahiri and Schmidt(1978)]
- [Prucha(1987)] showed that the standard errors produced by the GLS estimator of a triangular SUR model are not consistent

Estimate recursive model by sureg

```
. sureg (y1 = x1 x2) (y2 = y1 x2 x3) (y3 = y2 x3 x4) , isure nolog  tol(1e-15)
Seemingly unrelated regression, iterated
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
y1	500	2	1.728764	0.7246	1530.52	0.0000
y2	500	3	1.971366	0.8247	2387.49	0.0000
y3	500	3	1.887788	0.8561	2919.81	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
y1						
x1	.992947	.0374362	26.52	0.000	.9195735	1.066321
x2	.5402264	.0405265	13.33	0.000	.460796	.6196568
_cons	.8546342	.0775078	11.03	0.000	.7027217	1.006547
y2						
y1	.5160286	.0317023	16.28	0.000	.4538932	.5781639
x2	.5097058	.05344	9.54	0.000	.4049654	.6144462
x3	1.009926	.0420154	24.04	0.000	.9275772	1.092275
_cons	-1.02735	.0932221	-11.02	0.000	-1.210061	-.8446377
y3						
y2	.5732566	.0240356	23.85	0.000	.5261477	.6203655
x3	.2917947	.0509012	5.73	0.000	.1920302	.3915593
x4	.8197978	.0419108	19.56	0.000	.7376541	.9019415
_cons	.8690175	.086074	10.10	0.000	.7003156	1.037719

```
. estimates store sur_sureg
```

Comparing the results

```
. estimates table sur_sem sur_sureg, b se(%7.6g) keep(y1: y2: y3:)
```

Variable	sur_sem	sur_sureg
y1		
x1	.99294697	.99294699
	.03886	.03744
x2	.54022642	.54022642
	.04176	.04053
_cons	.85463424	.85463424
	.07752	.07751
y2		
y1	.51602852	.51602858
	.04638	.0317
x2	.50970588	.50970583
	.06272	.05344
x3	1.009926	1.009926
	.04299	.04202
_cons	-1.0273495	-1.0273495
	.09831	.09322
y3		
y2	.57325655	.57325658
	.04542	.02404
x3	.29179478	.29179474
	.07292	.0509
x4	.81979779	.81979779
	.04448	.04191
_cons	.86901749	.86901751
	.08962	.08607

legend: b/se

Path diagram
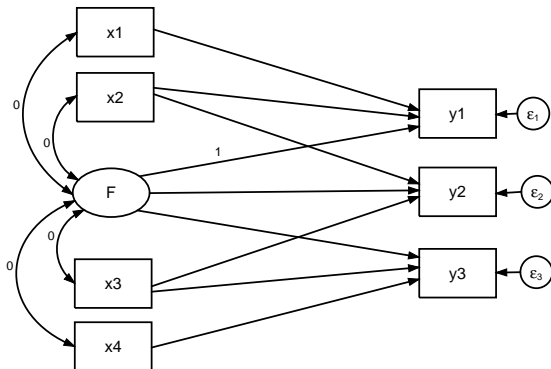
- SUR with observed exogenous variables and a latent variable

Mathematical description of model

- SUR model with observed exogenous variables and a latent variable

$$y_1 = \beta_{10} + F + \beta_{11}x_1 + \beta_{12}x_2 + \epsilon_1$$
$$y_2 = \beta_{20} + \rho_2 F + \beta_{22}x_2 + \beta_{23}x_3 + \epsilon_2$$
$$y_3 = \beta_{30} + \rho_3 F + \beta_{33}x_2 + \beta_{34}x_4 + \epsilon_3$$

where $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3)'$, $\mathbf{E}[\epsilon] = (0, 0, 0)'$, and
$\mathbf{Var}[\epsilon] = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$, $\mathbf{E}[F] = 0$, and $\mathbf{Var}[F] = \sigma_F^2$.

- sem (y1 <- F x1 x2) (y2 <- F x2 x3) (y3 <- F x3 x4) ,
  cov(F*(x1 x2 x3 x4)@0)

```
. sem    (y1 <- F x1 x2) (y2 <- F x2 x3) (y3 <- F x3 x4) ,          ///
>        cov(F*(x1 x2 x3 x4)@0 ) nolog
Endogenous variables
Observed: y1 y2 y3
Exogenous variables
Observed: x1 x2 x3 x4
Latent:   F
Structural equation model                    Number of obs    =       1000
Estimation method = ml
Log likelihood    = -13388.953
 ( 1)   [y1]F = 1
 ( 2)   [cov(x1,F)]_cons = 0
 ( 3)   [cov(x2,F)]_cons = 0
 ( 4)   [cov(x3,F)]_cons = 0
 ( 5)   [cov(x4,F)]_cons = 0
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Structural						
y1 <-						
x1	.9833928	.0308434	31.88	0.000	.9229408	1.043845
x2	.4752493	.0325475	14.60	0.000	.4114573	.5390413
F	1	(constrained)				
_cons	.9836244	.0564412	17.43	0.000	.8730018	1.094247
y2 <-						
x2	.4352693	.0306857	14.18	0.000	.3751264	.4954122
x3	1.02039	.0278183	36.68	0.000	.9658674	1.074913
F	.8578341	.1012175	8.48	0.000	.6594514	1.056217
_cons	-1.010054	.053389	-18.92	0.000	-1.114695	-.905414
y3 <-						
x3	.2991783	.0281943	10.61	0.000	.2439186	.3544381
x4	.7973739	.030267	26.34	0.000	.7380516	.8566962
F	.6020631	.0701647	8.58	0.000	.4645429	.7395833
_cons	1.086239	.0521731	20.82	0.000	.9839821	1.188497
Variance						
e.y1	1.692668	.1851758			1.366002	2.097452
e.y2	1.751188	.1469865			1.485549	2.064327
e.y3	2.180155	.1151949			1.965674	2.418038
F	1.48224	.2073715			1.126762	1.949868
Covariance						
x1						
F	0	(constrained)				
x2						
F	0	(constrained)				

```
Variance
         e.y1      1.692668    .1851758                        1.366002    2.097452
         e.y2      1.751188    .1469865                        1.485549    2.064327
         e.y3      2.180155    .1151949                        1.965674    2.418038
            F       1.48224    .2073715                        1.126762    1.949868

Covariance
   x1
            F             0   (constrained)

   x2
            F             0   (constrained)

   x3
            F             0   (constrained)

   x4
            F             0   (constrained)

LR test of model vs. saturated: chi2(6)   =      7.07, Prob > chi2 = 0.3144
. estimates store sur_sem
```

Path diagram
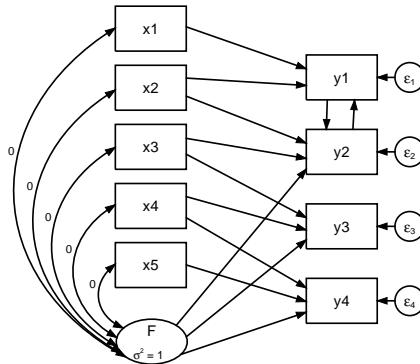
- Nonrecursive system with a latent variable

Mathematical description of model

- SUR model with observed exogenous variables and a latent variable

$$y_1 = \beta_{10} + \gamma_{12}y_2 + \beta_{11}x_1 + \beta_{12}x_2 + \epsilon_1$$
$$y_2 = \beta_{20} + \gamma_{21}y_1 + \rho_2 F + \beta_{22}x_2 + \beta_{23}x_3 + \epsilon_2$$
$$y_3 = \beta_{30} + \rho_3 F + \beta_{33}x_3 + \beta_{34}x_4 + \epsilon_3$$
$$y_4 = \beta_{40} + \rho_4 F + \beta_{44}x_4 + \beta_{45}x_5 + \epsilon_4$$

where $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4)'$, $\mathbf{E}\left[\epsilon\right] = (0, 0, 0, 0)'$,

$$\mathbf{Var}\left[\epsilon\right] = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix} \mathbf{E}\left[F\right] = 0, \text{ and } \mathbf{Var}\left[F\right] = 1.$$

- sem (y1 <- y2 x1 x2) (y2 <- y1 F x2 x3) (y3 <- F x3 x4)
  (y4 <- F x4 x5), cov(F*(x1 x2 x3 x4 x5)@0 F@1)

```
. sem  (y1 <- y2 x1 x2) (y2 <- y1 F x2 x3) (y3 <- F x3 x4)     ///
>      (y4 <- F x4 x5), cov(F*(x1 x2 x3 x4 x5)@0 F@1) nolog
Endogenous variables
Observed:  y1 y2 y3 y4
Exogenous variables
Observed:  x1 x2 x3 x4 x5
Latent:    F
Structural equation model                  Number of obs      =       1000
Estimation method = ml
Log likelihood     = -18510.86
 ( 1)  [cov(x1,F)]_cons = 0
 ( 2)  [cov(x2,F)]_cons = 0
 ( 3)  [cov(x3,F)]_cons = 0
 ( 4)  [cov(x4,F)]_cons = 0
 ( 5)  [cov(x5,F)]_cons = 0
 ( 6)  [var(F)]_cons = 1
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
Structural						
y1 <-						
y2	.7028302	.006432	109.27	0.000	.6902237	.7154368
x1	.969066	.0383729	25.25	0.000	.8938565	1.044276
x2	.4794069	.040371	11.88	0.000	.4002812	.5585327
_cons	1.017156	.0481089	21.14	0.000	.9228639	1.111447
y2 <-						
y1	.9965351	.0074365	134.01	0.000	.9819598	1.01111
x2	.5415499	.0421835	12.84	0.000	.4588718	.624228
x3	.9930632	.0393734	25.22	0.000	.9158927	1.070234
F	1.10419	.0884848	12.48	0.000	.9307631	1.277617
_cons	-.9047842	.0573361	-15.78	0.000	-1.017161	-.7924075
y3 <-						
x3	.2819725	.0294297	9.58	0.000	.2242913	.3396537
x4	.8366874	.0298951	27.99	0.000	.7780941	.8952806
F	.7132992	.0667452	10.69	0.000	.5824811	.8441174
_cons	1.112365	.0520723	21.36	0.000	1.010306	1.214425
y4 <-						
x4	1.00644	.0343128	29.33	0.000	.939188	1.073692
x5	.7611961	.0331479	22.96	0.000	.6962274	.8261648
F	1.233119	.0947673	13.01	0.000	1.047379	1.41886
_cons	1.131975	.0614426	18.42	0.000	1.01155	1.252401
Variance						
e.y1	2.298255	.1422471			2.035703	2.59467
e.y2	1.961872	.1884104			1.625266	2.368191
e.y3	2.195999	.1171369			1.978009	2.438014
e.y4	2.240033	.2153591			1.855321	2.704517

```
Variance
       e.y1 |   2.298255    .1422471                    2.035703     2.59467
       e.y2 |   1.961872    .1884104                    1.625266    2.368191
       e.y3 |   2.195999    .1171369                    1.978009    2.438014
       e.y4 |   2.240033    .2153591                    1.855321    2.704517
          F |          1  (constrained)
```

```
Covariance
  x1
          F |          0  (constrained)
```

```
  x2
          F |          0  (constrained)
```

```
  x3
          F |          0  (constrained)
```

```
  x4
          F |          0  (constrained)
```

```
  x5
          F |          0  (constrained)
```

```
LR test of model vs. saturated: chi2(13) =     12.47, Prob > chi2 = 0.4899
```

Standard postestimation

- Most standard postestimation features in Stata work after `sem`
  - `test`, `lrtest`, `lincom`, `testnl`, `nlcom`, `predict`, and the `estimates` commands are some important postestimation commands that work after `sem`

- `margins` does not work after `sem` because of the latent variables

Special postestimation

- Some of the important postestimation commands written or modified specifically for sem
  - estat gof, estat mindicies, estat scoretests, estat stdize, estat stable, and estat teffects

Direct and indirect effects

- `estat teffects` computes direct effect, indirect effects, total effects and their standard errors
- The direct effect of a variable x on an endogenous variable y is the coefficient on x in the equation for y
    - What is the change in y attributable to a unit change in x, conditional on all other variables in the equation
    - This effect ignores any simultaneous effects
- The total effect of a variable x is the change in an endogenous variable y attributable to a unit change in x after accounting for all the simultaneity in the system
    - Solve the system for the reduced form
    - The total effects are the coefficients in the reduced form specification
- The indirect effect of a variable is the total effect minus the direct effect

Direct effects example

```
. estat teffects, noindirect nototal
Direct effects
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
Structural						
y1 <-						
y1	0	(no path)				
y2	.7028302	.006432	109.27	0.000	.6902237	.7154368
x1	.969066	.0383729	25.25	0.000	.8938565	1.044276
x2	.4794069	.040371	11.88	0.000	.4002812	.5585327
x3	0	(no path)				
F	0	(no path)				
y2 <-						
y1	.9965351	.0074365	134.01	0.000	.9819598	1.01111
y2	0	(no path)				
x1	0	(no path)				
x2	.5415499	.0421835	12.84	0.000	.4588718	.624228
x3	.9930632	.0393734	25.22	0.000	.9158927	1.070234
F	1.10419	.0884848	12.48	0.000	.9307631	1.277617
y3 <-						
x3	.2819725	.0294297	9.58	0.000	.2242913	.3396537
x4	.8366874	.0298951	27.99	0.000	.7780941	.8952806
F	.7132992	.0667452	10.69	0.000	.5824811	.8441174
y4 <-						
x4	1.00644	.0343128	29.33	0.000	.939188	1.073692
x5	.7611961	.0331479	22.96	0.000	.6962274	.8261648
F	1.233119	.0947673	13.01	0.000	1.047379	1.41886

Total effects example

```
. estat teffects, noindirect nodirect
Total effects
```

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
Structural						
y1 <-						
y1	2.337728	.0174451	134.01	0.000	2.303537	2.37192
y2	2.345856	.0214684	109.27	0.000	2.303779	2.387934
x1	3.234479	.1133693	28.53	0.000	3.012279	3.456679
x2	2.870528	.1209785	23.73	0.000	2.633415	3.107642
x3	2.329583	.1075459	21.66	0.000	2.118798	2.540369
F	2.590271	.2172053	11.93	0.000	2.164557	3.015986
y2 <-						
y1	3.326164	.0248212	134.01	0.000	3.277515	3.374812
y2	2.337728	.021394	109.27	0.000	2.295797	2.37966
x1	3.223272	.1283112	25.12	0.000	2.971787	3.474757
x2	3.402132	.1416455	24.02	0.000	3.124512	3.679752
x3	3.314575	.1326908	24.98	0.000	3.054506	3.574644
F	3.685487	.2980858	12.36	0.000	3.101249	4.269724
y3 <-						
x3	.2819725	.0294297	9.58	0.000	.2242913	.3396537
x4	.8366874	.0298951	27.99	0.000	.7780941	.8952806
F	.7132992	.0667452	10.69	0.000	.5824811	.8441174
y4 <-						
x4	1.00644	.0343128	29.33	0.000	.939188	1.073692
x5	.7611961	.0331479	22.96	0.000	.6962274	.8261648
F	1.233119	.0947673	13.01	0.000	1.047379	1.41886

Conclusion

- SEM devotees know that I have only scratched the surface
- Simultaneous-equation types may be interested in including latent variables in their models
  - The postestimation commands, particularly `teffects`, can estimate partial effect parameters and compute specification tests that are not available from other commands for estimating the parameters of simultaneous equation models
  - Even if you are not interested in SEM, you may be interested in `sem`

📄 Lahiri, Kajal and Peter Schmidt. 1978. "A Note on the Consistency of the GLS Estimator in Triangular Structural Systems," *Econometrica*, 46, 1217–1221.

📄 Prucha, Ingmar R. 1987. "The Variance-Covariance Matrix of the Maximum Likelihood Estimator in Triangular Structural Systems: Consistent Estimation," *Econometrica*, 55, 977–978.