Generalizing sem in Stata

Jeff Pitblado
Director of Statistical Software
StataCorp LP

2013 Stata Conference
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 \texttt{gsem} command

- Think \textbf{G}eneralized \textbf{S}tructural \textbf{E}quations \textbf{M}odel
- Inspired by \texttt{gllamm} and \texttt{sem}
- Documented in [SEM]

Features

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

#### Syntax

```plaintext
gsem y <- x, family(family) link(link)
```

### Supported family and link combinations

<table>
<thead>
<tr>
<th></th>
<th>logit</th>
<th>probit</th>
<th>cloglog</th>
<th>log</th>
<th>identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>binomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ordinal</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multinomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>negative binomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>gamma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Gaussian</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

J. Pitblado (StataCorp)
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
  - $Lname[level_spec]#\ldots$

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
SEM Builder Demonstration
Example: oprobit for panel data

\begin{align*}
gsem y1 \leftarrow x z \ I[id1], \ family(\text{ordinal}) \ link(\text{probit}) \\
xtoprobit y1 x z
\end{align*}
Example: oprobit for multi-level data

```
gsem y1 <- x z I[id1] I[id1>id2], oprobit
meoprobit y1 x z || id1: || id2:
```

J. Pitblado (StataCorp)  Generalizing sem in Stata  2013 Stata Conference
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 l[id1], oprobit) (y2 <- x z l[id1], probit)
Example: bivariate, with 2 independent random effects

\[
gsem (y_1 \leftarrow x \ z \ I[id_1], \text{oprobit}) \ (y_2 \leftarrow x \ z \ J[id_1], \text{probit}), \text{covariance}(I[id_1]*J[id_1]@0)
\]
Example: bivariate, with 2 random effects

\[ \text{gsem (} y_1 \leftarrow x \ z \ I[id1], \text{oprobit)} (y_2 \leftarrow x \ z \ J[id1], \text{probit)} \]
Conclusion

What’s next

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

J. Pitblado (StataCorp)