Estimating a GLM with log link and Poisson regression for continuous variables

Tihomir Asparouhov

October 2, 2013

Generalized linear models (GLM) with the log link function are useful in modeling continuous positive outcomes. Suppose that we have a positive dependent variable Y and a predictor variable X and we want to estimate a model where

$$log(E(Y|X)) = \alpha + \beta X$$

or equivalently

$$E(Y|X) = Exp(\alpha + \beta X)$$

The last equation is also referred to as a Poisson regression and can be defined for continuous dependent variable, not just for Poisson/count variables. In Mplus this can be accomplished by estimating the model

$$Y \sim N(Exp(\alpha + \beta X), \sigma^2)$$

with the maximum-likelihood estimator. Thus we fit a normal distribution to the dependent variable Y with mean $Exp(\alpha + \beta X)$ and variance σ^2. In structural equation models the above Poisson regression can be used with latent variables as well. The dependent variable can be a latent variable, however the predictor variable has to be observed.

To estimate the above model in Mplus one has to use a special feature that allows the use of variables in the model constraint statements. A sample Mplus input statement that accomplishes this is as follows

```
data: file is 1.dat;
variable: names=x y;
constraint=x;
usevar=y;
model: [y] (mean);
```

model constraint: new(a b); mean=exp(a+b*x);

In the above input file the first row specifies the data file. The second row specifies the variables names. It also specifies that the variable Y will be the only variable used in the model. The variable X will be used in the parameter constraints model specified in MODEL CONSTRAINT. In the third row we give a name for the mean parameter of Y: mean. This name is used in MODEL CONSTRAINT in the fourth row where the actual exponential relationship between the covariate X and the mean of Y is specified. In this row we also introduce new model parameters which are the α and β parameters in the Poisson regression. More than one X variable can be handled, using more entries in the CONSTRAINT= statement.

Example 5.23 in the Mplus User's guide also illustrates the use of variables in the MODEL CONSTRAINT statements, however in that example the variance rather than the mean is modeled in the MODEL CONSTRAINT statement.

1 A Reciprocal Interaction Example

Consider the more complex model:

$$y2 = y1 + x1 + x2,$$
$$y1 = y2 + x1.$$

where the first equation is a linear regression and the second equation is a Poisson regression with log link and coefficients a, b1, and b2.

This can be analyzed as:

variable:
names=y1 y2 x1 x2;
usevar=y1 x2 y2d x1d;
constraint = y2 x1;
define: y2d=y2; x1d=x1;
model: [y1] (mean); y2d on y1 x1d x2;
model constraint: new(a b1 b2); mean=exp(a+b1*y2+b2*x1);

2 A Residual Covariance Example

Consider a model with a residual covariance:

$$y2 = x1 + x2$$
$$y1 = x1$$

where y1 is poisson with log link and y2 is gaussian, with identity link, and where there is a residual covariance between y1 and y2.
This can be analyzed as:

```
variable:
names=y1 y2 x1 x2;
usevar=y1 y2 x2 x1d;
constraint=x1;
define: x1d=x1;
model: [y1] (mean); y2 on x1d x2; y1 with y2;
model constraint: new(a b1); mean=exp(a+b1*x1);
```