Interaction and non-linear models

In recent years, there has been considerable interest in extending and applying structural equation models to situations where there are non-linear relationships involving latent variables, in particular to models with interaction effects. Various chapters in Schumacker & Marcoulides (1998) discuss the issues involved in non-linear structural equation models and describe different methods of estimation. The full information methods proposed by Jöreskog & Yang (1996) are difficult to apply in practice due to the complicated non-linear constraints that must be specified and the necessity to have large samples and use asymptotic covariance matrices. Bollen (1995, 1996) and Bollen & Paxton (1998) showed that non-linear models can be estimated easily with TSLS. Bollen & Paxton (1998) describe a rather complicated procedure to do this using SAS. Here we show how TSLS can be used easily and effectively using a single command in PRELIS.

To describe the basic idea of TSLS we use the Kenny–Judd model as a prototype for non-linear models (see Kenny & Judd, 1984). Bollen & Paxton (1998) demonstrated how TSLS can be used with several other types of nonlinear models.

1. Example: The Kenny–Judd Model

The Kenny–Judd model is

$$y = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta, \qquad (1)$$

where ξ_1 and ξ_2 are latent variables and ζ is a random error term assumed to be uncorrelated with ξ_1 and ξ_2. Kenny & Judd (1984) considered the case when there are two observable indicators x_1 and x_2 of ξ_1 and two observable indicators x_3 and x_4 of ξ_2, such that

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_2 & 0 \\ 0 & 1 \\ 0 & \lambda_4 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}. \qquad (2)$$

Kenny & Judd (1984) did not include the constant intercept terms α and τ_i in (1) and (2), but as argued by Jöreskog & Yang (1996), these are necessary for a correct analysis of the model.

Solving for ξ_1 and ξ_2 in the first and third equation in (2) and substituting into (1) gives

$$y_1 = \alpha + \gamma_1 x_1 + \gamma_2 x_3 + \gamma_3 x_1 x_3 + u, \qquad (3)$$

where

$$u = -\gamma_1\delta_1 - \gamma_2\delta_3 - \gamma_3\left(x_1\delta_3 + x_3\delta_1 - \delta_1\delta_3\right) + \zeta. \qquad (4)$$

Note that the γ 's in (3) are the same as in (1) but the error term *u* is not the same as ζ . The error term *u* in (3) is correlated with the variables on the right side in (3) so that ordinary least squares cannot be used to estimate this equation. However, as shown by Bollen & Paxton (1998), x_2, x_4, and x_2x_4 can be used as instrumental variables. If the model holds, these instrumental variables are uncorrelated with *u*. Assuming that raw data on *y*, x_1, x_2, x_3, and x_4, are available in the file **KJUDD.RAW**, the following PRELIS command file computes the necessary product variables and estimates the non-linear equation (3) directly (see file **KJTSLS1.PRL**):

```
Estimating Kenny-Judd Model by Bollen's TSLS
DA NI=5
LA
Y X1 X2 X3 X4
RA=KJUDD.RAW
CO ALL
NE X1X3=X1*X3
NE X1X4=X1*X4
NE X2X3=X2*X3
NE X2X4=X2*X4
RG Y ON X1 X3 X1X3 WITH X2 X4 X2X4 RES=U
OU RA=KJRES.RAW
```

The estimated equation is

```
       Y = 0.936 + 0.340*X1 + 0.399*X3 + 0.965*X1X3 + Error, R² = 0.594
Standerr  (1.011) (0.115)    (0.0883)   (0.164)
t-values   0.926   2.948      4.516      5.899
P-values   0.355   0.003      0.000      0.000
```

The raw data on all the variables, including the residual *u*, may be saved in a file by adding RA = KJRES.RAW on the OU command. Once this has been done, one can analyze these data and verify that indeed *u* is independent of x_2, x_4, and x_2x_4 but not independent of x_1, x_3, and x_1x_3. This is seen by running the following PRELIS command file (see **KJTSLS2.PRL**):

```
Checking the Residual
DA NI=10
LA
Y X1 X2 X3 X4 X1X3 X1X4 X2X3 X2X4 U
RA=KJRES.RAW
CO ALL
SD Y X1X4 X2X3
RG U ON X1 X3 X1X3
RG U ON X2 X4 X2X4
OU MA=KM
```

2. Estimation by Means of Latent Variable Scores

In this section, we illustrate yet another simple way of estimating the nonlinear equation (1) in the Kenny–Judd model, namely by means of latent variable scores.

The LSF file corresponding to **KJUDD.RAW** is **KJUDD.LSF**. This can be obtained by running the following PRELIS command file (see file: **KJUD2.PRL**):

```
Computing LSF file from KJUDD.RAW
DA NI=5
LA
Y X1 X2 X3 X4
RA=KJUDD.RAW
CO ALL
OU MA=CM RA=KJUDD.LSF
```

This run will also produce a DSF file called **KJUDD.DSF**. To obtain the latent variable scores for ξ_1 and ξ_2, use the data system file **KJUDD.DSF** and the following SIMPLIS command file (see file **KENJUDD.SPL**):

```
Estimating the Measurement Model in the Kenny-Judd Model and Latent Variable Scores
System File from File KJUDD.DSF
Latent Variables Ksi1 Ksi2
Relationships
X1=1*Ksi1
X2=Ksi1
X3=1*Ksi2
X4=Ksi2
LSFfile KJUDD.LSF
End of Problem
```

Alternatively, one can use the following LISREL command file (see file **KENJUDD.LIS**):

```
Estimating the Measurement Model in the Kenny-Judd Model
SY=KJUDD.LSF
SE
2 3 4 5 /
MO NX=4 NK=2 LX=FU,FI PH=SY,FR TD=DI,FR
LK
Ksi1 Ksi2
FR LX(2,1) LX(4,2)
VA 1 LX(1,1) LX(3,2)
LS=KJUDD.LSF
PD
OU ME=ML
```

Verify that the LSF file KJUDD.LSF has been appended with the scores on ξ_1 and ξ_2. One can now estimate (1) directly using the following PRELIS syntax file (see file **KENJUD3.PRL**):

```
Estimating Kenny-Judd Model from Latent Variable Scores
SY=KJUDD.LSF
CO ALL
NE  X1X2 = X1*X2
RG  Y  ON  X1 X2 X1X2
OU
```

The equation is estimated as

```
        Y = 1.135  + 0.192*X1 + 0.0905*X2 + 0.121*X1X2 + Error, R² = 0.102
 Standerr  (0.0251) (0.0255)   (0.0279)    (0.0239)
 t-values   45.265   7.528      3.244       5.041
 P-values   0.000    0.000      0.001       0.000
```

Error Variance = 0.570

The following chi-squares test the hypothesis that all regression coefficients are zero except the intercept.

Variable	-2lnL	Chi-square	df	Covariates
Y	2272.368	107.844	3	X1 X2 X1X2

Analysis of Variance Table

Regression	d.f.	Residual	d.f.	F	Covariates
64.689	3	568.071	996	37.806	X1 X2 X1X2