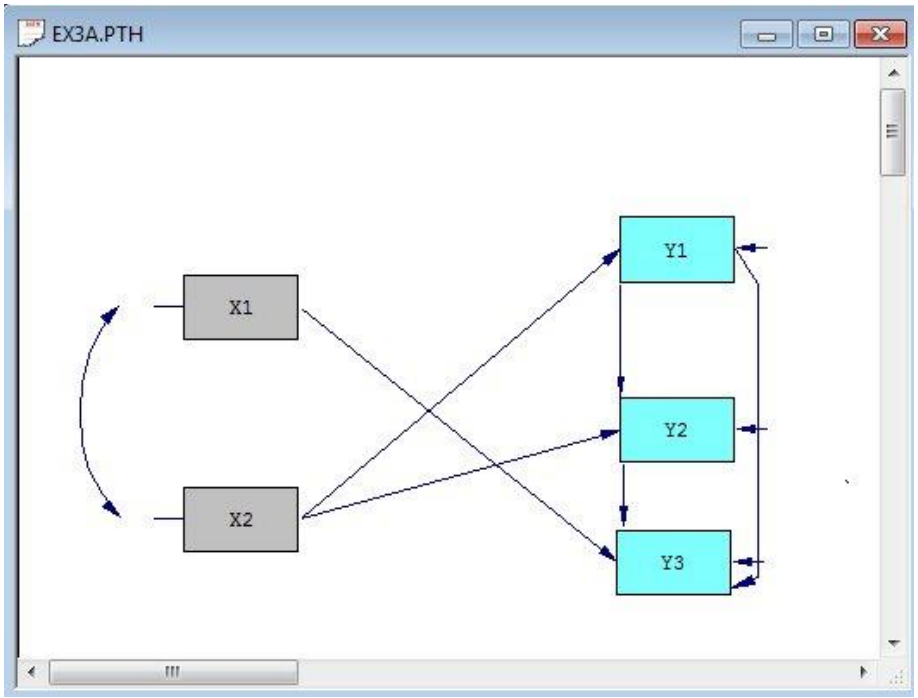
Path diagram: Union sentiment of textile workers

Path analysis, due to Wright (1934), is a technique to assess the direct causal contribution of one variable to another in a nonexperimental situation. The problem, in general, is that of estimating the coefficients of a set of linear *Structural equations*, representing the cause and effect relationships hypothesized by the investigator. The system involves variables of two kinds: independent or cause variables $x_1, x_2, \ldots, x_q$ and dependent or effect variables $y_1, y_2, \ldots, y_p$. The classical technique consists of first solving the structural equations for the dependent variables in terms of the independent variables and the random disturbance terms $z_1, z_2, \ldots, z_p$ to obtain the *reduced form equations*, estimating the regression of the dependent variables on the independent variables and then solving for the structural parameters in terms of the regression coefficients. The last step is not always possible. Models of this kind and a variety of estimation techniques have been extensively studied by econometricians (see Theil, 1971); by biometricians (see Turner and Stevens, 1959, and references therein); and by sociologists (see Blalock, 1985, and Duncan, 1975). Some of these models involve latent variables; see Duncan (2966), Werts & Linn (1970), and Hauser & Goldberger (1971).

In this example we shall consider a path analysis model for directly observed variables. Estimating a path analysis model for directly observed variables with LISREL is straightforward. Rather than estimating each equation separately, LISREL considers the model as a system of equations and estimates all the structural coefficients directly. The reduced form is obtained as a by-product.

The fundamental difference between this type of model and a regression model is that dependent variables appear also on the right side of the relationships. The following example is a *recursive system* in the sense that the dependent variables can be ordered in a sequence such that each dependent variable depends only on *x*-variables and previous dependent variables.

McDonald & Clelland (1984) analysed data on union sentiment of southern nonunion textile workers. After transformation of one variable and treatment of outliers, Bollen (1989a) reanalyzed a subset of the variables according to the model shown in the figure below.



The variables are:

y_1 = deference (submissiveness) to managers

y_2 = support for labor activism

y_3 = sentiment towards unions

x_1 = logarithm of years in textile mill

x_2 = age

The model shown above can be easily specified as relationships (equations):

```
Relationships
   Y1 = X2
   Y2 = X2 Y1
   Y3 = X1 Y1 Y2
```

These three lines express the following three statements, respectively:

- Y1 depends on X2
- Y2 depends on X2 and Y1
- Y3 depends on X1, Y1, and Y2

It is also easy to formulate the model by specifying its paths:

```
Paths
X1 -> Y3
X2 -> Y1 Y2
Y1 -> Y2 Y3
Y2 -> Y3
```

These lines express the following statements:

- There is one path from X1 to X3
- There are two paths from X2, one to Y1 and one to Y2
- There are two paths from Y1, one to Y2 and one to Y3
- There is one path from Y2 to Y3.

Altogether there are six paths in the model corresponding to the six one-way (unidirected) arrows in the figure.

Table: Covariance matrix for union sentiment variables

y_1	y_2	y_3	x_1	x_2
14.610				
-5.250	11.017			
-8.057	11.087	31.971		
-0.482	0.677	1.559	1.021	
-18.857	17.861	28.250	7.139	215.662

The input file for this problem is straightforward (**EX3A.SPL** in the **Simplis Examples** folder):

EX3A.SPL

```
Title
   Union Sentiment of Textile Workers
   Variables: Y1 = deference (submissiveness) to managers
              Y2 = support for labor activism
              Y3 = sentiment towards unions
              X1 = years in textile mill
              X2 = age

Observed Variables: Y1 - Y3 X1 X2
Covariance matrix:
   14.610
   -5.250  11.017
   -8.057  11.087  31.971
   -0.482   0.677   1.559   1.021
  -18.857  17.861  28.250   7.139 215.662
Sample Size 173
Relationships
   Y1 = X2
   Y2 = X2 Y1
   Y3 = X1 Y1 Y2
Path Diagram
End of problem
```

The first nine lines are title lines that define the problem and the variables. To indicate the beginning of the title lines, one may use the word `Title`, although this is not necessary. The first real command line begins with the words `Observed Variables`. Note that Y1, Y2 and Y3 may be labeled collectively as Y1-Y3. The observed variables, the covariance matrix, and the sample size are defined as in the example Regression on GNP (**EX1A.SPL**). The model is specified as relationships.

The model makes a fundamental distinction between two kinds of variables: dependent and independent. The *dependent variables* are supposed to be explained by the model, i.e., variation and covariation in the independent variables are supposed to be accounted for by the *independent variables*. The dependent variables are on the left side of the equal sign. They correspond to those variables in the path diagram which have one-way arrows pointing towards them. Variables which are not dependent are called independent. The distinction between dependent and independent variables is already inherent in the labels: *x*-variables are independent, and *y*-variables are dependent. Note that *y*-variables can appear on the right side of equations.

The output gives the following estimated relationships:

```
       Structural Equations


       Y1 =  - 0.0874*X2, Errorvar.= 12.961, R² = 0.113
 Standerr     (0.0187)             (1.402)
 Z-values     -4.664                9.247
 P-values      0.000                0.000

       Y2 =  - 0.285*Y1 + 0.0579*X2, Errorvar.= 8.488 , R² = 0.230
 Standerr     (0.0619)   (0.0161)             (0.918)
 Z-values     -4.598      3.597                9.247
 P-values      0.000      0.000                0.000

       Y3 =  - 0.218*Y1 + 0.850*Y2 + 0.861*X1, Errorvar.= 19.454, R² = 0.390
 Standerr     (0.0974)   (0.112)    (0.341)             (2.104)
 Z-values     -2.235      7.555      2.526               9.247
 P-values      0.025      0.000      0.012               0.000
```