Example 10 — MIMIC model

Description   Remarks and examples   Reference   Also see

Description

To demonstrate a MIMIC model, we use the following summary statistics data:

```
. use https://www.stata-press.com/data/r19/sem_mimic1
(Multiple indicators and multiple causes)

. ssd describe

  Summary statistics data from
  https://www.stata-press.com/data/r19/sem_mimic1.dta
  Observations:             432          Multiple indicators and multip..
     Variables:               5          31 Oct 2024 08:29
                                         (_dta has notes)
  ------------------------------------------------------------------------
  Variable name                  Variable label
  ------------------------------------------------------------------------
  occpres                        Occupational prestige, two-digit Dunca..
  income                         Total family income in units of $2000,..
  s_occpres                      Subjective occupational prestige
  s_income                       Subjective income
  s_socstat                      Subjective overall social status
  ------------------------------------------------------------------------

. notes

_dta:
  1.  Summary statistics data from Kluegel, J. R., R. Singleton, Jr., and C. E.
      Starnes. 1977. Subjective class identification: A multiple indicator
      approach. American Sociological Review 42: 599-611.
      https://doi.org/10.2307/2094558.
  2.  Data are also analyzed in Bollen, K. A. 1989. Structural Equations with
      Latent Variables. New York: Wiley.
  3.  The summary statistics represent 432 white adults included in the sample
      for the 1969 Gary Area Project for the Institute of Social Research at
      Indiana University.
  4.  The three subjective variables are measures of socioeconomic status based
      on an individuals perception of their own income, occupational prestige,
      and social status.
  5.  The income and occpres variables are objective measures of income and
      occupational prestige, respectively.
```

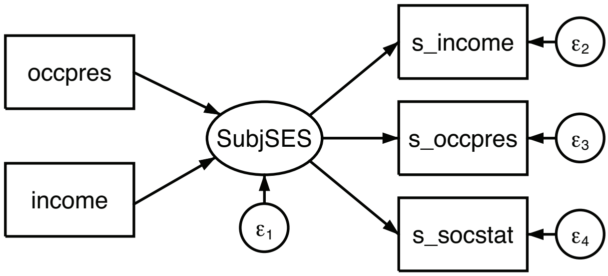
See *Structural models 10: MIMIC models* in [SEM] **Intro 5** for background.

Remarks and examples

Remarks are presented under the following headings:

Fitting the MIMIC model

Based on the data referenced above, Bollen (1989, 397–399) fits a MIMIC model, the path diagram of which is



Bollen includes paths that he constrains and we do not show. Our model is nonetheless equivalent to the one he shows. In his textbook, Bollen illustrates various ways the same model can be written.

```
. sem (SubjSES -> s_income s_occpres s_socstat) (SubjSES <- income occpres)
Endogenous variables
  Measurement: s_income s_occpres s_socstat
  Latent:      SubjSES

Exogenous variables
  Observed: income occpres

Fitting target model:
Iteration 0:  Log likelihood = -4252.1834  (not concave)
Iteration 1:  Log likelihood = -4022.8854  (not concave)
Iteration 2:  Log likelihood = -3994.3867
Iteration 3:  Log likelihood = -3978.7751  (not concave)
Iteration 4:  Log likelihood = -3974.6636
Iteration 5:  Log likelihood = -3972.0269
Iteration 6:  Log likelihood = -3971.9238
Iteration 7:  Log likelihood = -3971.9236

Structural equation model                            Number of obs = 432
Estimation method: ml

Log likelihood = -3971.9236

 ( 1)  [s_income]SubjSES = 1

------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  SubjSES    |
      income |    .082732   .0138498     5.97   0.000     .0555869    .1098772
     occpres |   .0046275   .0012464     3.71   0.000     .0021847    .0070704
-------------+----------------------------------------------------------------
Measurement  |
  s_income   |
     SubjSES |          1  (constrained)
       _cons |   .9612091   .0794151    12.10   0.000     .8055583     1.11686
  -----------+----------------------------------------------------------------
  s_occpres  |
     SubjSES |   .7301352   .0832915     8.77   0.000     .5668869    .8933835
       _cons |   1.114563   .0656195    16.99   0.000     .9859512    1.243175
  -----------+----------------------------------------------------------------
  s_socstat  |
     SubjSES |   .9405161   .0934855    10.06   0.000     .7572878    1.123744
       _cons |   1.002113   .0706576    14.18   0.000      .863627      1.1406
-------------+----------------------------------------------------------------
var(e.s_in~e)|   .2087546   .0254098                      .1644474    .2649996
var(e.s_oc~s)|   .2811852   .0228914                      .2397153    .3298291
var(e.s_so~t)|   .1807129   .0218405                      .1425987    .2290146
var(e.Subj~S)|   .1860097   .0270476                      .1398822    .2473481
------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(4) = 26.65           Prob > chi2 = 0.0000
```

Notes:

1. In this model, there are three observed variables that record the respondents' ideas of their perceived socioeconomic status (SES). One is the respondent's general idea of his or her SES (`s_socstat`); another is based on the respondent's income (`s_income`); and the last is based on the respondent's occupational prestige (`s_occpres`). Those three variables form the latent variable `SubjSES`.

2. The other two observed variables are the respondent's income (`income`) and occupation, the latter measured by the two-digit Duncan SEI scores for occupations (`occpres`). These two variables are treated as predictors of `SubjSES`.

3. In the model, item 1 is viewed as subjective and item 2 is viewed as objective.
4. All variables are statistically significant at the 5% level, but the model versus saturated test suggests that we are not modeling the covariances well.

Fitting the MIMIC model with the Builder

Use the diagram above for reference.

1. Open the dataset.

   In the Command window, type

   ```
   . use https://www.stata-press.com/data/r19/sem_mimic1
   ```

2. Open a new Builder diagram.

   Select menu item **Statistics > SEM (structural equation modeling) > Model building and estimation**.

3. Create the measurement component for subject socioeconomic status.

   Select the Add measurement component tool, , and then click near the center of the diagram.

   In the resulting dialog box,

   a. change the *Latent variable name* to `SubjSES`;
   b. select `s_income`, `s_occpress`, and `s_socstat` by using the *Measurement variables* control;
   c. select `Right` in the *Measurement direction* control;
   d. click on **OK**.

   If you wish, move this component by clicking on any variable and dragging it.

4. Create the two variables for the formative indicators of socioeconomic status.

   a. Select the Add observed variable tool, □, and then click in the diagram to add the new variable slightly above and to the left of `SubjSES`. After adding it, you can click inside the rectangle and move the variable if you wish.
   b. In the Contextual Toolbar, select `occpress` by using the *Variable* control.
   c. Continuing with the □ tool, click in the diagram to add another new variable below the `occpress` variable.
   d. In the Contextual Toolbar, select `income` by using the *Variable* control.

5. Create the paths for the formative indicators of socioeconomic status.

   a. Select the Add path tool, →.
   b. Click in the lower-right quadrant of the `occpress` rectangle (it will highlight when you hover over it), and drag a path to the upper-left quadrant of the `SubjSES` oval (it will highlight when you can release to connect the path).
   c. Continuing with the → tool, click in the upper-right quadrant of the `income` rectangle, and drag a path to the lower-left quadrant of the `SubjSES` oval.

6. Clean up the direction of the errors.

   The error on SubjSES is likely to have been created atop one of the existing paths. Choose the Select tool, , and then click in the SubjSES oval. Click on one of the **Error rotation** buttons, , in the Contextual Toolbar until the error is where you want it.

7. Estimate.

   Click on the **Estimate** button, , in the Standard Toolbar, and then click on **OK** in the resulting *SEM estimation options* dialog box.

Tip: See the tips of [SEM] **Example 9** to make creating paths somewhat easier than described above.

You can open a completed diagram in the Builder by typing

```
. webgetsem sem_mimic1
```

Evaluating the residuals with estat residuals

Remember that SEM fits covariances and means. Residuals in the SEM sense thus refer to covariances and means. If we are not fitting well, we can examine the residuals.

```
. estat residuals, normalized
Residuals of observed variables
  Mean residuals
```

	s_income	s_occpres	s_socstat	income	occpres
raw	0.000	0.000	0.000	0.000	0.000
normalized	0.000	0.000	0.000	0.000	0.000

```
  Covariance residuals
```

	s_income	s_occpres	s_socstat	income	occpres
s_income	0.000				
s_occpres	-0.009	0.000			
s_socstat	0.000	0.008	0.000		
income	0.101	-0.079	-0.053	0.000	
occpres	-0.856	1.482	0.049	0.000	0.000

```
  Normalized covariance residuals
```

	s_income	s_occpres	s_socstat	income	occpres
s_income	0.000				
s_occpres	-0.425	0.000			
s_socstat	0.008	0.401	0.000		
income	1.362	-1.137	-0.771	0.000	
occpres	-1.221	2.234	0.074	0.000	0.000

Notes:

1. The residuals can be partitioned into two subsets: mean residuals and covariance residuals.
2. The normalized option caused the normalized residuals to be displayed.
3. Concerning mean residuals, the raw residuals and the normalized residuals are shown on a separate line of the first table.

4. Concerning covariance residuals, the raw residuals and the normalized residuals are shown in separate tables.
5. Distinguish between normalized residuals and standardized residuals. Both are available from `estat residuals`; if we wanted standardized residuals, we would have specified the `standardized` option instead of or along with `normalized`.
6. Both normalized and standardized residuals attempt to adjust the residuals in the same way. The normalized residuals are always valid, but they do not follow a standard normal distribution. The standardized residuals do follow a standard normal distribution if they can be calculated; otherwise, they will equal missing values. When both can be calculated (equivalent to both being appropriate), the normalized residuals will be a little smaller than the standardized residuals.
7. The normalized covariance residuals between `income` and `s_income` and between `occpres` and `s_occpres` are large.

Performing likelihood-ratio tests with lrtest

Thus Bollen (1989, 397–399) suggests adding a direct path from the objective measures to the corresponding subjective measures. We are about to fit the model

```
(SubjSES -> s_income s_occpres s_socstat)   ///
(SubjSES <- income occpres)                 ///
(s_income <- income)                        ///  <- new
(s_occpres <- occpres)                      //   <- new
```

For no other reason than we want to demonstrate the likelihood-ratio test, we will then use `lrtest` rather than `test` to test the joint significance of the new paths. `lrtest` compares the likelihood values of two fitted models. Thus we will use `lrtest` to compare this new model with the one above. To do that, we must plan ahead and store in memory the currently fit model:

```
. estimates store mimic1
```

Alternatively, we could skip that and calculate the joint significance of the two new paths by using a Wald test and the `test` command.

In any case, having stored the current estimates under the name `mimic1`, we can now fit our new model:

```
. sem (SubjSES -> s_income s_occpres s_socstat)
>     (SubjSES <- income occpres)
>     (s_income <- income)
>     (s_occpres <- occpres)
Endogenous variables
  Observed:    s_income s_occpres
  Measurement: s_socstat
  Latent:      SubjSES
Exogenous variables
  Observed: income occpres
Fitting target model:
Iteration 0:  Log likelihood = -4267.0974  (not concave)
Iteration 1:  Log likelihood = -4022.6745  (not concave)
Iteration 2:  Log likelihood = -3977.0648
Iteration 3:  Log likelihood = -3962.9229
Iteration 4:  Log likelihood = -3962.1604
Iteration 5:  Log likelihood = -3960.8404
Iteration 6:  Log likelihood = -3960.7133
Iteration 7:  Log likelihood = -3960.7111
Iteration 8:  Log likelihood = -3960.7111
Structural equation model                           Number of obs = 432
Estimation method: ml
Log likelihood = -3960.7111
 ( 1)  [s_income]SubjSES = 1
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  s_income   |
     SubjSES |          1  (constrained)
      income |   .0532425   .0142862     3.73   0.000      .025242     .081243
       _cons |   .8825314   .0781685    11.29   0.000     .7293239    1.035739
  -----------+----------------------------------------------------------------
  s_occpres  |
     SubjSES |    .783781   .1011457     7.75   0.000      .585539     .982023
     occpres |   .0045201   .0013552     3.34   0.001     .0018641    .0071762
       _cons |    1.06586   .0696058    15.31   0.000     .9294353    1.202285
  -----------+----------------------------------------------------------------
  SubjSES    |
      income |   .0538025   .0129158     4.17   0.000      .028488    .0791171
     occpres |   .0034324   .0011217     3.06   0.002     .0012339    .0056309
-------------+----------------------------------------------------------------
Measurement  |
  s_socstat  |
     SubjSES |   1.195539   .1582735     7.55   0.000     .8853282    1.505749
       _cons |    1.07922    .078323    13.78   0.000     .9257097     1.23273
-------------+----------------------------------------------------------------
var(e.s_in~e)|   .2292697   .0248905                      .1853261    .2836329
var(e.s_oc~s)|   .2773786   .0223972                      .2367783    .3249407
var(e.s_so~t)|   .1459009    .028228                      .0998556    .2131785
var(e.Subj~S)|   .1480275   .0278381                      .1023918    .2140029
------------------------------------------------------------------------------
LR test of model vs. saturated: chi2(2) = 4.22            Prob > chi2 = 0.1211
```

Now we can perform the likelihood-ratio test:

```
. lrtest mimic1 .

Likelihood-ratio test
Assumption: mimic1 nested within .

 LR chi2(2) =  22.42
Prob > chi2 = 0.0000
```

Notes:

1. The syntax of `lrtest` is `lrtest` *modelname1 modelname2*. We specified the first model name as `mimic1`, the model we previously stored. We specified the second model name as period (`.`), meaning the model most recently fit. The order in which we specify the names is irrelevant.

2. We find the two added paths to be extremely significant.

Reference

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley. https://doi.org/10.1002/9781118619179.

Also see

[SEM] **Example 36g** — MIMIC model (generalized response)

[SEM] **sem** — Structural equation model estimation command

[SEM] **estat residuals** — Display mean and covariance residuals

[SEM] **lrtest** — Likelihood-ratio test of linear hypothesis