Example with Categorical Indicators

Mplus
WLSMV (Robust DWLS Approach) with Theta Parameterization
(note: results generated with Mplus 8.9)

```
INPUT INSTRUCTIONS


  title:  CESD second order model with categorical designation;

  data:  file=c:\jason\mplus\negex\wave1\pncesdw1.dat;
               format=35f1.0;
         listwise=on;

  variable:  names = w1adv1 w1adv2 w1adv3 w1hlp1 w1hlp2
                     w1hlp3 w1vst1 w1vst2 w1vst3 w1emo1
                     w1emo2 w1emo3 w1unw1 w1unw2 w1unw3
                     w1dwn1 w1dwn2 w1dwn3 w1out1 w1out2
                     w1out3 w1fai1 w1fai2 w1fai3 w1dboth
                     w1dblues w1dmind w1ddep w1defrt w1dsleep
                     w1dhappy w1dlone w1dnogo w1dsad w1denjoy;


            usevariable= w1dboth w1dblues w1dmind w1ddep w1defrt
               w1dsleep w1dhappy w1dsad w1denjoy;

            categorical=w1dboth w1dblues w1dmind w1ddep w1defrt
                     w1dsleep w1dhappy w1dsad w1denjoy;

  analysis:  type=general;  estimator=wlsmv; parameterization=theta;
      !WLSMV is the default and estimator= is not needed here;
      !parameterization=theta changes the default delta parameterization to theta;
      !WLSMV gives probit estimates;

  model: negaff by w1dboth w1dblues w1ddep w1dsad;
         somatic by w1dmind w1defrt w1dsleep;
         positive by w1dhappy  w1denjoy;

  output: stdyx;


INPUT READING TERMINATED NORMALLY



CESD second order model with categorical designation;

SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                         720

Number of dependent variables                                    9
Number of independent variables                                  0
Number of continuous latent variables                            3

Observed dependent variables

  Binary and ordered categorical (ordinal)
   W1DBOTH     W1DBLUES    W1DMIND     W1DDEP      W1DEFRT     W1DSLEEP
   W1DHAPPY    W1DSAD      W1DENJOY

Continuous latent variables
   NEGAFF      SOMATIC     POSITIVE


Estimator                                                    WLSMV
Maximum number of iterations                                  1000
Convergence criterion                                    0.500D-04
Maximum number of steepest descent iterations                   20
Parameterization                                             THETA
Link                                                        PROBIT
```

```
MODEL FIT INFORMATION

Number of Free Parameters                       39

Chi-Square Test of Model Fit

          Value                             39.254*
          Degrees of Freedom                    24
          P-Value                           0.0257

*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
    for chi-square difference testing in the regular way.  MLM, MLR and WLSM
    chi-square difference testing is described on the Mplus website.  MLMV, WLSMV,
    and ULSMV difference testing is done using the DIFFTEST option.

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.030
          90 Percent C.I.                    0.010  0.046
          Probability RMSEA <= .05           0.982

CFI/TLI

          CFI                                0.997
          TLI                                0.995

Chi-Square Test of Model Fit for the Baseline Model

          Value                           4645.035
          Degrees of Freedom                    36
          P-Value                           0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                              0.023

Optimum Function Value for Weighted Least-Squares Estimator

          Value                     0.12781725D-01


MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 NEGAFF   BY
    W1DBOTH            1.000      0.000    999.000    999.000
    W1DBLUES           2.327      0.290      8.018      0.000
    W1DDEP             3.097      0.382      8.104      0.000
    W1DSAD             2.775      0.328      8.473      0.000

 SOMATIC  BY
    W1DMIND            1.000      0.000    999.000    999.000
    W1DEFRT            1.521      0.219      6.937      0.000
    W1DSLEEP           0.758      0.108      7.026      0.000

 POSITIVE BY
    W1DHAPPY           1.000      0.000    999.000    999.000
    W1DENJOY           0.804      0.181      4.430      0.000

 SOMATIC  WITH
    NEGAFF             0.444      0.066      6.710      0.000

 POSITIVE WITH
    NEGAFF             0.980      0.176      5.567      0.000
    SOMATIC            0.896      0.174      5.144      0.000

  Variances
    NEGAFF             0.461      0.088      5.259      0.000
    SOMATIC            0.665      0.128      5.190      0.000
    POSITIVE           3.852      1.142      3.373      0.001


STANDARDIZED MODEL RESULTS
STDYX Standardization
```

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 NEGAFF   BY
    W1DBOTH            0.562      0.037     15.365      0.000
    W1DBLUES           0.845      0.023     36.571      0.000
    W1DDEP             0.903      0.015     61.117      0.000
    W1DSAD             0.883      0.015     58.328      0.000

 SOMATIC  BY
    W1DMIND            0.632      0.037     17.288      0.000
    W1DEFRT            0.779      0.032     24.155      0.000
    W1DSLEEP           0.526      0.041     12.673      0.000

 POSITIVE BY
    W1DHAPPY           0.891      0.027     32.725      0.000
    W1DENJOY           0.845      0.028     30.113      0.000

 SOMATIC  WITH
    NEGAFF             0.802      0.031     25.520      0.000

 POSITIVE WITH
    NEGAFF             0.736      0.029     25.538      0.000
    SOMATIC            0.560      0.048     11.590      0.000


R-SQUARE

    Observed                                        Two-Tailed     Scale
    Variable        Estimate       S.E.  Est./S.E.    P-Value     Factors

    W1DBOTH            0.315      0.041      7.682      0.000       0.827
    W1DBLUES           0.714      0.039     18.285      0.000       0.535
    W1DMIND            0.400      0.046      8.644      0.000       0.775
    W1DDEP             0.815      0.027     30.559      0.000       0.430
    W1DEFRT            0.606      0.050     12.078      0.000       0.628
    W1DSLEEP           0.277      0.044      6.337      0.000       0.851
    W1DHAPPY           0.794      0.049     16.363      0.000       0.454
    W1DSAD             0.780      0.027     29.164      0.000       0.469
    W1DENJOY           0.713      0.047     15.057      0.000       0.535
```

lavaan
WLSMV (Robust DWLS Approach) with Theta Parameterization

```
> ## Categorical data analysis with lavaan
> 
> #because the data were fixed format, special statements are needed
> pncesdw1= read.fortran("c:/jason/mplus/negex/wave1/pncesdw1.dat",
+          c("F1.0","F1.0","F1.0","F1.0","F1.0",
+         "F1.0","F1.0","F1.0","F1.0","F1.0",
+         "F1.0","F1.0","F1.0","F1.0","F1.0",
+         "F1.0","F1.0","F1.0","F1.0","F1.0",
+         "F1.0","F1.0","F1.0","F1.0","F1.0",
+         "F1.0","F1.0","F1.0","F1.0","F1.0",
+         "F1.0","F1.0","F1.0","F1.0","F1.0"))
> names(pncesdw1) = c("w1adv1","w1adv2","w1adv3","w1hlp1","w1hlp2",
+                "w1hlp3","w1vst1","w1vst2","w1vst3","w1emo1",
+                "w1emo2","w1emo3","w1unw1","w1unw2","w1unw3",
+                "w1dwn1","w1dwn2","w1dwn3","w1out1","w1out2",
+                "w1out3","w1fai1","w1fai2","w1fai3","w1dboth",
+                "w1dblues","w1dmind","w1ddep","w1defrt","w1dsleep",
+                "w1dhappy","w1dlone","w1dnogo","w1dsad","w1denjoy")
> 
> 
> twofact = '
+         negaff =~ w1dboth + w1dblues + w1ddep + w1dsad
+          somatic =~ w1dmind + w1defrt + w1dsleep
+          positive =~ w1dhappy  + w1denjoy
+ '
> 
> model1= sem(twofact, data = pncesdw1, missing = 'listwise',ordered=c("w1dboth",
+                "w1dblues","w1dmind","w1ddep","w1defrt","w1dsleep",
+                "w1dhappy","w1dlone","w1dnogo","w1dsad","w1denjoy"),estimator="wlsmv",
+                           parameterization="theta")
> summary(model1,fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)
```

(excerpts)

```
lavaan 0.6-18 ended normally after 86 iterations

  Estimator                                       DWLS
  Optimization method                           NLMINB
  Number of model parameters                        39

  Number of observations                           720

Model Test User Model:
                                              Standard      Scaled
  Test Statistic                                18.380      39.203
  Degrees of freedom                                24          24
  P-value (Chi-square)                           0.784       0.026
  Scaling correction factor                                  0.503
  Shift parameter                                            2.646
    simple second-order correction

Model Test Baseline Model:

  Test statistic                              8225.833    4638.605
  Degrees of freedom                                36          36
  P-value                                        0.000       0.000
  Scaling correction factor                                  1.779

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    1.000       0.997
  Tucker-Lewis Index (TLI)                       1.001       0.995

  Robust Comparative Fit Index (CFI)                         0.986
  Robust Tucker-Lewis Index (TLI)                            0.980

Root Mean Square Error of Approximation:

  RMSEA                                          0.000       0.030
  90 Percent confidence interval - lower         0.000       0.010
  90 Percent confidence interval - upper         0.021       0.046
  P-value H_0: RMSEA <= 0.050                    1.000       0.982
  P-value H_0: RMSEA >= 0.080                    0.000       0.000

  Robust RMSEA                                               0.053
  90 Percent confidence interval - lower                     0.025
  90 Percent confidence interval - upper                     0.078
  P-value H_0: Robust RMSEA <= 0.050                         0.402
  P-value H_0: Robust RMSEA >= 0.080                         0.038

Standardized Root Mean Square Residual:

  SRMR                                           0.029       0.029

Parameter Estimates:

  Parameterization                               Theta
  Standard errors                           Robust.sem
  Information                                 Expected
  Information saturated (h1) model        Unstructured

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  negaff =~
    w1dboth           1.000                                0.679    0.562
    w1dblues          2.327    0.290    8.012    0.000    1.579    0.845
    w1ddep            3.097    0.382    8.099    0.000    2.102    0.903
    w1dsad            2.776    0.328    8.467    0.000    1.884    0.883
  somatic =~
    w1dmind           1.000                                0.816    0.632
    w1defrt           1.520    0.219    6.933    0.000    1.240    0.779
    w1dsleep          0.758    0.108    7.021    0.000    0.618    0.526
  positive =~
    w1dhappy          1.000                                1.962    0.891
    w1denjoy          0.804    0.182    4.428    0.000    1.577    0.845

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  negaff ~~
    somatic           0.444    0.066    6.706    0.000    0.802    0.802
    positive          0.980    0.176    5.564    0.000    0.736    0.736
  somatic ~~
    positive          0.896    0.174    5.141    0.000    0.560    0.560

Thresholds:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
    w1dboth|t1        0.190    0.057    3.316    0.001    0.190    0.157
    w1dboth|t2        1.432    0.079   18.133    0.000    1.432    1.185
    w1dboth|t3        2.132    0.110   19.454    0.000    2.132    1.764
```

```
    w1dblues|t1      1.352    0.138    9.801    0.000    1.352    0.723
    w1dblues|t2      2.456    0.185   13.291    0.000    2.456    1.314
    w1dblues|t3      3.100    0.215   14.420    0.000    3.100    1.658
    w1ddep|t1        0.888    0.139    6.376    0.000    0.888    0.382
    w1ddep|t2        2.928    0.237   12.341    0.000    2.928    1.258
    w1ddep|t3        3.860    0.292   13.210    0.000    3.860    1.658
    w1dsad|t1        0.533    0.111    4.779    0.000    0.533    0.250
    w1dsad|t2        2.542    0.181   14.041    0.000    2.542    1.192
    w1dsad|t3        3.508    0.215   16.345    0.000    3.508    1.645
    w1dmind|t1       0.281    0.063    4.489    0.000    0.281    0.218
    w1dmind|t2       1.409    0.086   16.467    0.000    1.409    1.092
    w1dmind|t3       2.140    0.118   18.115    0.000    2.140    1.658
    w1defrt|t1       0.100    0.075    1.325    0.185    0.100    0.063
    w1defrt|t2       1.413    0.120   11.755    0.000    1.413    0.887
    w1defrt|t3       1.957    0.144   13.550    0.000    1.957    1.228
    w1dsleep|t1      0.049    0.055    0.892    0.372    0.049    0.042
    w1dsleep|t2      0.659    0.061   10.824    0.000    0.659    0.561
    w1dsleep|t3      1.164    0.071   16.503    0.000    1.164    0.990
    w1dhappy|t1      1.226    0.184    6.670    0.000    1.226    0.557
    w1dhappy|t2      2.672    0.327    8.172    0.000    2.672    1.213
    w1dhappy|t3      4.170    0.441    9.460    0.000    4.170    1.893
    w1denjoy|t1      1.472    0.159    9.285    0.000    1.472    0.788
    w1denjoy|t2      2.364    0.212   11.141    0.000    2.364    1.266
    w1denjoy|t3      3.536    0.268   13.174    0.000    3.536    1.893
```

Mplus
MLR—Robust Marginal Maximum Likelihood with Categorical Variables

Although unlikely to be of interest here for loadings, the path coefficients obtained with categorical ML estimation can be converted to odds ratios, using e^{β}, where e is the mathematical constant (approximately 2.72) and β is the unstandardized structural path coefficient. The robust adjustments (`estimator=MLR`) appear to do better than nonrobust marginal maximum likelihood (`estimator=ML;` Bandalos, 2014).

Because, I used the same model statements as above, I omitted them to save space. The analysis statements were changed to these, however.

```
analysis:  type=general;  estimator=mlr;
    !estimator=mlr invokes full maximum likelihood with robust adjustments
    ! when categorical variables are identified. Estimates are logit with this
    ! method and odds ratios can be used with predictive paths;
```

(output excerpts)

```
Estimator                                                         MLR
Information matrix                                           OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Integration Specifications
  Type                                                       STANDARD
  Number of integration points                                     15
  Dimensions of numerical integration                               3
  Adaptive quadrature                                              ON
Link                                                            LOGIT
Cholesky                                                           ON


THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION

Number of Free Parameters                        39

Loglikelihood

          H0 Value                        -5396.880
          H0 Scaling Correction Factor       1.0506
            for MLR

Information Criteria

          Akaike (AIC)                    10871.759
          Bayesian (BIC)                  11050.350
          Sample-Size Adjusted BIC        10926.514
            (n* = (n + 2) / 24)
```

```
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes**

          Pearson Chi-Square
          Value                          6871.310
          Degrees of Freedom               261972
          P-Value                          1.0000

          Likelihood Ratio Chi-Square
          Value                          1464.685
          Degrees of Freedom               261972
          P-Value                          1.0000

** Of the 262144 cells in the frequency table, 132
   were deleted in the calculation of chi-square due to extreme values.

MODEL RESULTS
                                                           Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 NEGAFF   BY
    W1DBOTH            1.000      0.000    999.000    999.000
    W1DBLUES           2.191      0.300      7.296      0.000
    W1DDEP             3.184      0.486      6.548      0.000
    W1DSAD             2.730      0.382      7.152      0.000

 SOMATIC  BY
    W1DMIND            1.000      0.000    999.000    999.000
    W1DEFRT            1.420      0.210      6.779      0.000
    W1DSLEEP           0.714      0.115      6.236      0.000

 POSITIVE BY
    W1DHAPPY           1.000      0.000    999.000    999.000
    W1DENJOY           0.939      0.208      4.502      0.000

 SOMATIC  WITH
    NEGAFF             1.438      0.234      6.149      0.000
 POSITIVE WITH
    NEGAFF             2.895      0.555      5.218      0.000
    SOMATIC            2.613      0.559      4.675      0.000

 Thresholds
    W1DBOTH$1          0.319      0.097      3.300      0.001
    W1DBOTH$2          2.486      0.153     16.287      0.000
    W1DBOTH$3          3.852      0.234     16.461      0.000
    W1DBLUES$1         2.226      0.229      9.738      0.000
    W1DBLUES$2         4.173      0.339     12.311      0.000
    W1DBLUES$3         5.429      0.442     12.273      0.000
    W1DMIND$1          0.487      0.108      4.499      0.000
    W1DMIND$2          2.449      0.169     14.464      0.000
    W1DMIND$3          3.826      0.235     16.305      0.000
    W1DDEP$1           1.548      0.245      6.332      0.000
    W1DDEP$2           5.232      0.517     10.125      0.000
    W1DDEP$3           7.243      0.665     10.887      0.000
    W1DEFRT$1          0.176      0.126      1.400      0.162
    W1DEFRT$2          2.427      0.220     11.010      0.000
    W1DEFRT$3          3.376      0.282     11.973      0.000
    W1DSLEEP$1         0.084      0.091      0.928      0.353
    W1DSLEEP$2         1.095      0.105     10.414      0.000
    W1DSLEEP$3         1.965      0.128     15.354      0.000
    W1DHAPPY$1         2.114      0.319      6.634      0.000
    W1DHAPPY$2         4.681      0.596      7.854      0.000
    W1DHAPPY$3         7.086      0.935      7.581      0.000
    W1DSAD$1           0.897      0.187      4.787      0.000
    W1DSAD$2           4.427      0.373     11.853      0.000
    W1DSAD$3           6.390      0.501     12.744      0.000
    W1DENJOY$1         2.842      0.361      7.880      0.000
    W1DENJOY$2         4.590      0.529      8.676      0.000
    W1DENJOY$3         6.740      0.793      8.498      0.000

 Variances
    NEGAFF             1.485      0.311      4.769      0.000
    SOMATIC            2.149      0.475      4.524      0.000
    POSITIVE          10.737      3.305      3.249      0.001
```

```
STANDARDIZED MODEL RESULTS

STDYX Standardization
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 NEGAFF   BY
    W1DBOTH            0.558      0.040     13.842      0.000
    W1DBLUES           0.827      0.028     29.557      0.000
    W1DDEP             0.906      0.018     50.302      0.000
    W1DSAD             0.878      0.019     45.509      0.000

 SOMATIC  BY
    W1DMIND            0.629      0.042     14.958      0.000
    W1DEFRT            0.754      0.038     20.055      0.000
    W1DSLEEP           0.500      0.044     11.371      0.000

 POSITIVE BY
    W1DHAPPY           0.875      0.032     27.708      0.000
    W1DENJOY           0.861      0.031     27.379      0.000

 SOMATIC  WITH
    NEGAFF             0.805      0.037     22.018      0.000

 POSITIVE WITH
    NEGAFF             0.725      0.042     17.440      0.000
    SOMATIC            0.544      0.061      8.947      0.000

R-SQUARE
    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value
    W1DBOTH            0.311      0.045      6.921      0.000
    W1DBLUES           0.684      0.046     14.779      0.000
    W1DMIND            0.395      0.053      7.479      0.000
    W1DDEP             0.821      0.033     25.151      0.000
    W1DEFRT            0.569      0.057     10.027      0.000
    W1DSLEEP           0.250      0.044      5.686      0.000
    W1DHAPPY           0.765      0.055     13.854      0.000
    W1DSAD             0.771      0.034     22.755      0.000
    W1DENJOY           0.742      0.054     13.690      0.000
```

lavaan

Marginal maximum likelihood estimation for binary and ordered categorical variables is not presently supported by lavaan except for some confirmatory factor models. Keep your eye on this page, http://lavaan.ugent.be/tutorial/cat.html, because maximum likelihood for categorical estimators may be added in the near future.

References

Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: a multidisciplinary journal*, *21*, 102-116.