Dealing with non-normality in xtdpdml

Last revised December 11, 2016

Richard Williams, Enrique Moral-Benito and Paul D. Allison
http://academicweb.nd.edu/~rwilliam/dynamic/index.html

By default, the `xtdpdml` model assumes that observed variables have a multivariate normal distribution. When the normality assumption holds Maximum Likelihood provides the smallest possible standard errors (i.e., it is efficient) but the SEs may be misleading under non-normality. How problematic violations of the normality assumption are depends on the circumstances. The ML estimator is consistent and asymptotically normal even under non-normality. Further, simulations by Moral-Benito et al (2016) show that under non-normal Data Generating Processes (DGPs) the ML estimator performs relatively well in finite samples compared to GMM. When the normality assumption is possibly problematic, the Stata `sem` command (and hence `xtdpdml`) provides various ways of relaxing the assumption. This note will explain three of the approaches and some of the advantages and disadvantages of each. The appendix includes examples that illustrate the points made here.

First, with Stata 14 and later, the `vce(sbentler)` option can be specified. As Stata Corp explains on its web pages (2016a),

> Stata's linear `sem` now provides the Satorra–Bentler scaled chi-squared test for model goodness of fit versus the saturated model… The likelihood-ratio test comparing your estimated model to the saturated model is derived under the assumption that the observed variables in your model are normally distributed. If they are not, that test is not appropriate. The Satorra–Bentler scaled chi-squared test is robust to nonnormality… [What's more] The same adjustment that gives you the Satorra–Bentler scaled chi-squared test makes a host of other things robust to nonnormality: standard errors, p-values, and confidence intervals reported by `sem` and standard errors, p-values, and confidence intervals for most posthoc comparisons and tests.

Note that `vce(sbentler)` relaxes the normality assumption when estimating standard errors but does NOT affect the coefficient estimates, i.e., regardless of whether you specify `vce(sbentler)` or not the coefficient estimates will be the same.

Unfortunately, a key limitation of the `vce(sbentler)` option is that it does NOT work with full-information maximum likelihood, i.e., it requires the use of listwise deletion. If missing data is a concern, researchers may prefer to use a different option, `vce(robust)`. As Stata Corp (2016a) also points out in the same on-line document,

> Stata's `sem` already had an adjustment that makes everything in "What's more" true. It is often called the Huber or White method, or just called the linearized estimator. Whatever you call it, this estimator and the Satorra–Bentler adjustment are making your inferences robust to similar things. They are derived and computed differently, so they produce different estimates. As samples become very large, however, they converge to the same estimates.

When `vce(robust)` is specified, along with the default Maximum Likelihood estimation method, Stata calls the estimation method quasi-maximum likelihood (QML). Like `vce(sbentler)`, QML relaxes the normality assumption when estimating standard errors but does not affect the coefficient estimates, i.e., regardless of whether you specify `vce(robust)` or not the coefficient estimates will be the same.

A key advantage of `vce(robust)` is that, unlike `vce(sbentler)`, it can be used with Full Information Maximum Likelihood, i.e., it does NOT require listwise deletion of missing data. Since the standard errors from `vce(robust)` and `vce(sbentler)` are asymptotically equivalent, `vce(robust)` may be preferred when missing data are a concern. However, unlike `vce(sbentler)`, `vce(robust)` does not provide many goodness of fit measures. A possible strategy might be to specify the model without using `vce(robust)`, use goodness of fit measures to identify ways in which the model could be improved (e.g. relax the constraint that the effects of the *x*s are invariant across time), and then re-estimate the model using `vce(robust)`. On the other hand, if the assumption of normality is violated, the use of FIML may also be problematic.

With both `vce(robust)` and `vce(sbentler)`, specifying the option changes the standard errors but the coefficient estimates remain the same. A third approach is the asymptotic distribution free (ADF) estimation method, which is achieved by specifying `method(adf)`. As the Stata 14.2 (2016b, p. 44) manual explains,

> ADF makes no assumption of joint normality or even symmetry, whether for observed or latent variables. Whereas QML handles nonnormality by adjusting standard errors and not point estimates, ADF produces justifiable point estimates and standard errors under nonnormality… Be aware, however, that ADF is less efficient than ML when latent variables can be assumed to be normally distributed. If latent variables (including errors) are not normally distributed, on the other hand, ADF will produce more efficient estimates than ML or QML.

Like `vce(sbentler)`, ADF requires listwise deletion of missing data, which could be a major disadvantage in some cases. Also, `method(adf)` does not work with `technique(bhhh)`. Since `technique(nr 25 bhhh 25)` is currently the default in `xtdpdml`, the technique option will also have to be specified if you specify `method(adf)`. Our own very limited tests also suggest that models using ADF are harder to estimate and more likely to have convergence problems.

In conclusion, the ideal situation is when multivariate normality can be safely assumed. When this is not a safe assumption, at least three different approaches can be used, but each has advantages and disadvantages. In particular, `vce(sbentler)` and `method(adf)` do not permit the use of FIML, which may be undesirable if there is a lot of missing data. On the other hand, FIML also assumes multivariate normality, which makes its use questionable when that is not true. Ergo, if you do not feel comfortable assuming multivariate normality but want to use FIML, `vce(robust)` may be the best choice. But, if you do not feel comfortable with FIML either (or do not have much missing data) `vce(sbentler)` may be the best option because of the goodness of fit measures that it provides that `vce(robust)` does not.[1]

---

[1] As a sidelight, while the Stata 14 `sem` manual provides dozens of examples, in only one of them (Example one, Single factor measurement model) is any of these three approaches (`vce(sbentler)`) used. Rightly or wrongly, adjustments for non-normality may not be that common with analyses of structural equation models.

References

Moral-Benito, Enrique, Paul Allison and Richard Williams. 2016 (in progress; last revised October 29, 2016). "Dynamic Panel Data Modeling using Maximum Likelihood: An Alternative to Arellano-Bond". http://academicweb.nd.edu/~rwilliam/dynamic/Benito_Allison_Williams.pdf

Stata Corporation. 2016a. *Satorra–Bentler adjustments*. http://www.stata.com/new-in-stata/sem-satorra-bentler/. Last accessed December 2, 2016.

Stata Corporation. 2016b. *Stata Structural Equation Modeling Reference Manual Release 14*. Stata Press: College Station, Texas 77845.

Appendix: Examples

The following examples illustrate the default approach that assumes multivariate normality, followed by each of the three approaches that relaxes that assumption.

```
. use http://academicweb.nd.edu/~rwilliam/statafiles/wages, clear
. * Default approach assuming multivariate normality
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) gof

Highlights: Baseline Model
------------------------------------------------------------------------------
             |                 OIM
         wks |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wks          |
         wks |
         L1. |   .1871266   .0201939     9.27   0.000     .1475473    .2267059
             |
       lwage |
         L1. |   .6417917   .4842304     1.33   0.185    -.3072823    1.590866
             |
       union |
         L1. |  -1.191349   .5168951    -2.30   0.021    -2.204445   -.1782536
             |
          ed |  -.1122267   .0559477    -2.01   0.045    -.2218822   -.0025711
------------------------------------------------------------------------------
# of units = 595. # of periods = 7. First dependent variable is from period 2.
LR test of model vs. saturated: chi2(71)  =     110.23, Prob > chi2 =  0.0020
IC Measures: BIC =   25470.43  AIC =   24772.64
Wald test of all coeff = 0: chi2(4) =      90.09, Prob > chi2 =  0.0000
```

```
----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Likelihood ratio     |
         chi2_ms(71) |    110.228   model vs. saturated
            p > chi2 |      0.002
         chi2_bs(99) |   1059.393   baseline vs. saturated
            p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
               RMSEA |      0.030   Root mean squared error of approximation
 90% CI, lower bound |      0.019
         upper bound |      0.041
              pclose |      0.999   Probability RMSEA <= 0.05
---------------------+------------------------------------------------------
Information criteria |
                 AIC |  24772.644   Akaike's information criterion
                 BIC |  25470.425   Bayesian information criterion
---------------------+------------------------------------------------------
Baseline comparison  |
                 CFI |      0.959   Comparative fit index
                 TLI |      0.943   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
                SRMR |      0.022   Standardized root mean squared residual
                  CD |      0.313   Coefficient of determination
----------------------------------------------------------------------------
```

```
. * Now use vce(sbentler). Coefficients stay the same.
. * Standard errors and GOF measures change
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) vce(sbentler) gof
```

```
Highlights: Baseline Model
------------------------------------------------------------------------------
             |           Satorra-Bentler
         wks |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wks          |
         wks |
         L1. |   .1871266   .0387666     4.83   0.000     .1111453    .2631078
             |
       lwage |
         L1. |   .6417917   .6367676     1.01   0.314    -.6062498    1.889833
             |
       union |
         L1. |  -1.191349   .8912287    -1.34   0.181    -2.938125    .5554269
             |
          ed |  -.1122267   .0747721    -1.50   0.133    -.2587773     .034324
------------------------------------------------------------------------------
# of units = 595. # of periods = 7. First dependent variable is from period 2.
LR test of model vs. saturated: chi2(71)  =     110.23, Prob > chi2 =  0.0020
IC Measures: BIC =   25470.43  AIC =   24772.64
Wald test of all coeff = 0: chi2(4) =      28.32, Prob > chi2 =  0.0000
```

```
----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Likelihood ratio     |
         chi2_ms(71) |    110.228   model vs. saturated
            p > chi2 |      0.002
         chi2_bs(99) |   1059.393   baseline vs. saturated
            p > chi2 |      0.000
                     |
  Satorra-Bentler    |
       chi2sb_ms(71) |     63.675
            p > chi2 |      0.719
       chi2sb_bs(99) |    632.409
            p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
               RMSEA |      0.030   Root mean squared error of approximation
 90% CI, lower bound |      0.019
         upper bound |      0.041
              pclose |      0.999   Probability RMSEA <= 0.05
                     |
  Satorra-Bentler    |
            RMSEA_SB |      0.000   Root mean squared error of approximation
---------------------+------------------------------------------------------
Information criteria |
                 AIC |  24772.644   Akaike's information criterion
                 BIC |  25470.425   Bayesian information criterion
---------------------+------------------------------------------------------
Baseline comparison  |
                 CFI |      0.959   Comparative fit index
                 TLI |      0.943   Tucker-Lewis index
                     |
  Satorra-Bentler    |
              CFI_SB |      1.000   Comparative fit index
              TLI_SB |      1.019   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
                SRMR |      0.022   Standardized root mean squared residual
                  CD |      0.313   Coefficient of determination
----------------------------------------------------------------------------

. * vce(sbentler) does NOT work with fiml
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) vce(sbentler) fiml gof
vce(sbentler) not allowed with method(mlmv)
r(198);
```

```
. * Now use vce(robust). Coefficients stay the same, standard errors change.
. * In these particular examples vce(sbentler) and vce(robust) produce very
. * similar estimates of the standard errors.
. * But, few GOF measures are reported with vce(robust).
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) vce(robust) gof

Highlights: Baseline Model
------------------------------------------------------------------------------
             |               Robust
         wks |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wks          |
         wks |
         L1. |   .1871266   .0393371     4.76   0.000     .1100273    .2642258
             |
       lwage |
         L1. |   .6417917    .617815     1.04   0.299    -.5691035    1.852687
             |
       union |
         L1. |  -1.191349   .9082205    -1.31   0.190    -2.971429    .5887302
             |
          ed |  -.1122267   .0745572    -1.51   0.132    -.2583561    .0339027
------------------------------------------------------------------------------
# of units = 595. # of periods = 7. First dependent variable is from period 2.
Warning: LR test of model vs saturated could not be computed
IC Measures: BIC =   25457.65  AIC =   24768.64
Wald test of all coeff = 0: chi2(4) =      28.04, Prob > chi2 =  0.0000

----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Size of residuals    |
                SRMR |      0.022   Standardized root mean squared residual
                  CD |      0.313   Coefficient of determination
----------------------------------------------------------------------------
Note: model was fit with vce(robust); only stats(residuals) valid.

. * vce(robust) does work with fiml
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) vce(robust) fiml gof
```

[Output not shown. Since there is no missing data the results are identical to vce(robust) without fiml].

```
. * Now use method(adf). Both coefficients and standard errors change.
. * But, won't converge and gives few GOF measures
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) method(adf)
technique(nr) gof

Highlights: Baseline Model
------------------------------------------------------------------------------
             |                   .
         wks |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wks          |
         wks |
         L1. |   .3305837   .0059357    55.69   0.000     .3189499    .3422175
             |
       lwage |
         L1. |   .5552284    .001805   307.60   0.000     .5516906    .5587662
             |
       union |
         L1. |  -1.122687   .0039246  -286.06   0.000    -1.130379   -1.114995
             |
          ed |   .3869419    .001548   249.96   0.000     .3839078    .3899759
------------------------------------------------------------------------------
# of units = 595. # of periods = 7. First dependent variable is from period 2.
Warning: LR test of model vs saturated could not be computed
Warning: IC Measures BIC and AIC could not be computed
Wald test of all coeff = 0: chi2(1) =    3101.81, Prob > chi2 =  0.0000
Warning! Convergence not achieved

----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Discrepancy          |
          chi2_ms(.) |          .   model vs. saturated
            p > chi2 |          .
         chi2_bs(99) |    272.218   baseline vs. saturated
            p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
               RMSEA |          .   Root mean squared error of approximation
 90% CI, lower bound |      0.000
         upper bound |          .
              pclose |          .   Probability RMSEA <= 0.05
---------------------+------------------------------------------------------
Baseline comparison  |
                 CFI |      1.000   Comparative fit index
                 TLI |          .   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
                SRMR |  1.095e+08   Standardized root mean squared residual
                  CD |      0.248   Coefficient of determination
----------------------------------------------------------------------------

. * Also fiml will not work with adf
. xtdpdml wks L.lwage, inv(ed) pre(L.union) ti(Baseline Model) method(adf)
technique(nr) gof fiml
You cannot specify both fiml and method(adf)
Job is terminating.
```