SOCY7708: Hierarchical Linear Modeling
Instructor: Natasha Sarkisian
Class notes: HLM Diagnostics

Like OLS, HLM models rely on certain assumptions that have to be satisfied in order for regression coefficients to be unbiased and efficient estimates of the parameters of interest. Therefore, it is important to watch out for possible assumption violations and to take steps to prevent them. We will address the issues of model specification, collinearity, homoscedasticity, normality of level 1 and level 2 residuals, and linearity.

1. Model specification.

In HLM models, the issue of model specification concerns two main questions: (1) Did we include the right fixed effects? (2) Did we include the right random components? As we discussed, when specifying your model, you should rely heavily on your theory as well as utilize hypothesis testing. But there are some additional steps you can take to prevent model misspecification:

- Consider including aggregates of level 1 variables. It is always possible that what appears to be an effect of a level 1 variable is, in reality, an effect of its level 2 aggregate. The only way to test is to introduce such an aggregate. So far, we discussed aggregates to the mean, but sometimes, it is also possible to use group-level standard deviations. For example, you can use MEANSES to indicate the average level of SES in the school and DEVSES (within-school standard deviation) to indicate how diverse each school is in terms of SES. Such diversity may have an impact above and beyond the impact of the average level.

```
. bysort id: egen devses=sd(ses)

. mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses
i.female##i.sector devses || id: female, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(9)      =     861.54
Log likelihood = -23217.242                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.852182    .147225    19.37   0.000     2.563626    3.140738
     meanses |   3.123813   .5007382     6.24   0.000     2.142384    4.105242
             |
       c.ses#|
   c.meanses |   .7762981   .2691847     2.88   0.004     .2487057     1.30389
             |
         ses |          0  (omitted)
    1.sector |    1.00463   .4108992     2.44   0.014     .1992828    1.809978
             |
sector#c.ses |
```

```
         1  |   -1.54621   .2220969    -6.96   0.000    -1.981512   -1.110908
            |
    1.female |  -1.218688   .2383675    -5.11   0.000     -1.68588   -.7514966
     meanses |          0  (omitted)
            |
     female#|
   c.meanses |
         1  |  -.0401053   .5063098    -0.08   0.937    -1.032454    .9522437
            |
     female#|
     sector |
       1 1  |   .0546438   .4182702     0.13   0.896    -.7651507    .8744384
            |
      devses |  -2.666202   1.457351    -1.83   0.067    -5.522557    .1901533
       _cons |   14.57225   1.041436    13.99   0.000     12.53107    16.61343
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                var(female) |   .9479572   .5671689      .2934333    3.062444
                 var(_cons) |   2.934085   .5636358      2.013527     4.27551
           cov(female,_cons) |  -1.170931   .4832245     -2.118034   -.2238285
-----------------------------+------------------------------------------------
              var(Residual) |    36.3806   .6189185      35.18755    37.61411
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 183.93                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

- Consider including level 2 predictors of level 1 slopes (i.e., cross-level interactions) if you find significant variation in these slopes
- If the proportion of explained variance (R-squared – we discussed earlier how to calculate R-squared within each level and overall) is substantially reduced when you add a fixed effect, that can be a sign of misspecification.
- Sometimes a fixed effect misspecification (e.g., a nonlinearity) can lead to a misspecification of the random effects (excluded curvilinear effect may show up as a significant variance component for the slope). We will return to the issue of linearity below.

To prevent the misspecification problems in terms of random components:

- Always test whether each of your level 1 slopes varies across level 2 units (i.e., try to estimate each slope as random). However, you have to be careful not to "overtax" your data – if you have very few lower level cases within each upper level unit, you can't do too many random components.
- If the model doesn't converge or if it takes a long time to converge, that may mean that the model has too many random effects and the data are relatively sparse. In general, you should be cautious in specifying level-1 coefficients as random – as the number of random effects grows, the number of variances/covariances to be estimated increases even faster (for m random predictors, there are 1+m(m+1)/2 variance covariance components). As the number of random effects grows, significantly mode information is required to obtain reasonable estimates of variance/covariance components. The maximum depends on a number of factors: the magnitude of the variance components,

the degree of intercorrelation among the random effects, the magnitude of sigma squared, and other characteristics of the data.

- If there are high correlations among level-1 coefficients (i.e., slopes for different variables—correlations with the intercept are ok), the model must be simplified. There are a number of ways of dealing with it. You can, for example, use factor analysis to form scales and reduce the number of variables. You can also constrain one or more random effects to be zero (i.e. keep only the fixed effect for that variable), thus eliminating the correlation. This works well if that random effect is in fact negligible in size and/or non-significant.

2. Multicollinearity

Like regular OLS, HLM models can be affected by multicollinearity. There are no tools to check for it specifically for mixed command, but you can check basic correlations among your independent variables as well as variance inflation factors (VIFs) for the same model estimated with OLS:

```
. pwcorr  mathach ses female meanses sector size

             |  mathach      ses   female  meanses   sector     size
-------------+------------------------------------------------------
     mathach |   1.0000
         ses |   0.3608   1.0000
      female |  -0.1231  -0.0679   1.0000
     meanses |   0.3437   0.5306  -0.0589   1.0000
      sector |   0.2040   0.1896   0.0065   0.3573   1.0000
        size |  -0.0506  -0.0673  -0.0388  -0.1268  -0.4237   1.0000

. reg mathach ses female meanses sector size

      Source |       SS       df       MS              Number of obs =    7185
-------------+------------------------------           F(  5,  7179) =  315.35
       Model |  61205.6611     5  12241.1322           Prob > F      =  0.0000
    Residual |  278671.273  7179  38.8175614           R-squared     =  0.1801
-------------+------------------------------           Adj R-squared =  0.1795
       Total |  339876.934  7184  47.3102637           Root MSE      =  6.2304

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.148034   .1113801    19.29   0.000     1.929697    2.366372
      female |  -1.321295   .1478042    -8.94   0.000    -1.611034   -1.031555
     meanses |   2.889622   .2206451    13.10   0.000     2.457093    3.322151
      sector |   1.503238   .1724585     8.72   0.000     1.165169    1.841308
        size |   .0003457   .0001345     2.57   0.010     .0000821    .0006093
       _cons |   12.32108   .2222729    55.43   0.000     11.88536     12.7568
------------------------------------------------------------------------------

. vif
    Variable |       VIF       1/VIF
-------------+----------------------
     meanses |      1.54    0.648947
         ses |      1.39    0.717093
      sector |      1.38    0.726733
        size |      1.22    0.818586
      female |      1.01    0.992362
-------------+----------------------
    Mean VIF |      1.31
```

Different researchers advocate for different cutoff points for VIF. Some say that if any one of VIF values is larger than 4, there are some multicollinearity problems associated with that variable. Others use cutoffs of 5 or even 10. It can also be useful to check level 2 separately using means of your dependent variable as an outcome:

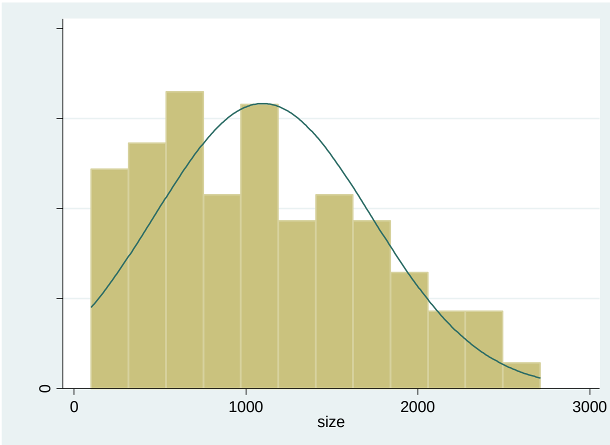
```
. bysort id: egen mathachm=mean(mathach)

. reg mathachm meanses sector size if tagged==1

      Source |       SS           df       MS      Number of obs   =       160
-------------+----------------------------------   F(3, 156)       =     99.84
       Model |  1016.16465         3   338.72155   Prob > F        =    0.0000
    Residual |  529.275622       156  3.39279245   R-squared       =    0.6575
-------------+----------------------------------   Adj R-squared   =    0.6509
       Total |  1545.44027       159  9.71975014   Root MSE        =     1.842

------------------------------------------------------------------------------
    mathachm | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     meanses |   5.359654   .3777613    14.19   0.000     4.613467    6.105841
      sector |   1.524517    .349298     4.36   0.000     .8345531    2.214481
        size |   .0005152   .0002603     1.98   0.050     1.00e-06    .0010293
       _cons |   11.38922   .4055502    28.08   0.000     10.58814     12.1903
------------------------------------------------------------------------------

. vif

    Variable |       VIF       1/VIF  
-------------+----------------------
      sector |      1.42    0.706227
        size |      1.26    0.794707
     meanses |      1.15    0.872531
-------------+----------------------
    Mean VIF |      1.27
```

When running your mixed models, you can also watch out for potential signs of multicollinearity (e.g., large coefficients for two correlated variables going in opposite directions, high standard errors).

3. Normality

HLM models assume that the level-1 and level 2 error terms are normally distributed. To make sure this assumption will be met, it is important to do some preliminary data screening before running mixed models. It is especially important to ensure that your dependent variable distribution is as close to normal as possible, but you should check independent variables as well. If substantial deviations from normality are identified, consider fixing them with a transformation. Note that when examining normality of level 2 variables, you should either have a separate level 2 file or you should limit your analysis to one record per higher level unit.

```
. egen tagged=tag(id)

. histogram size if tagged==1
(bin=12, start=100, width=217.75)
```

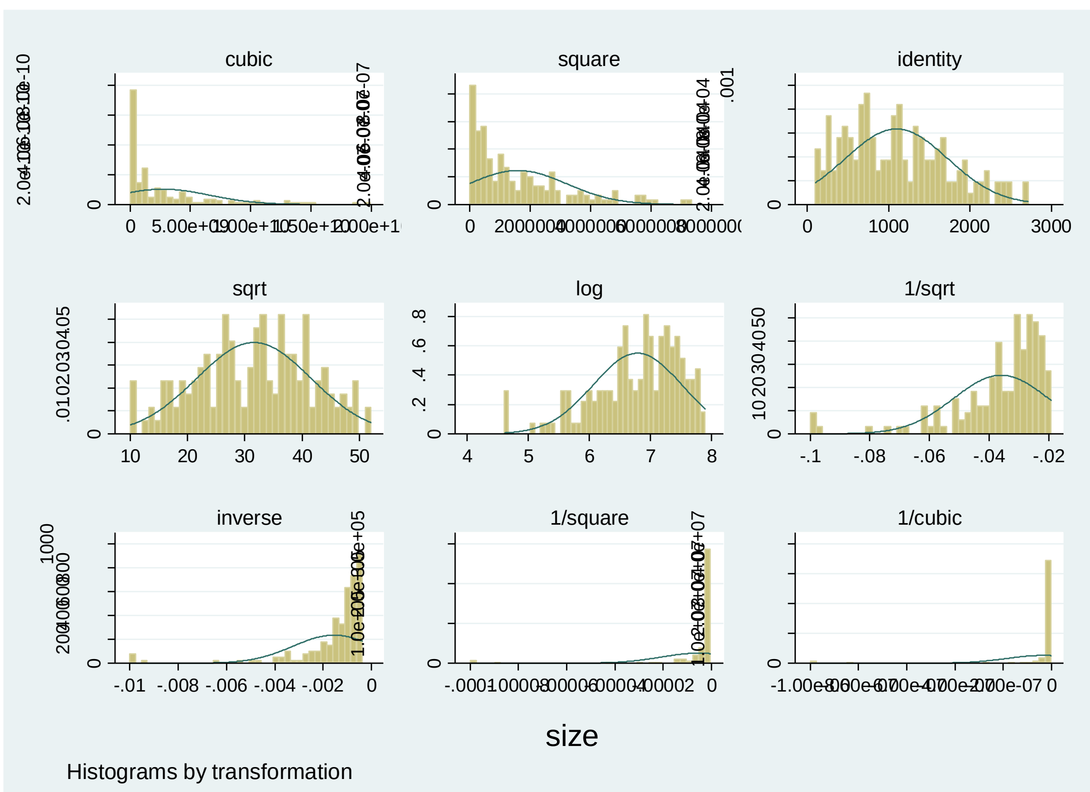


Looks like a right skew; to find a transformation:

```
. ladder size if tagged==1
```

Transformation	formula	chi2(2)	P(chi2)
cubic	size^3	60.02	0.000
square	size^2	31.36	0.000
identity	size	8.37	0.015
square root	sqrt(size)	7.18	0.028
log	log(size)	16.55	0.000
1/(square root)	1/sqrt(size)	58.10	0.000
inverse	1/size	.	0.000
1/square	1/(size^2)	.	0.000
1/cubic	1/(size^3)	.	0.000

```
. gladder size if tagged==1
```



Square root looks the best, so we would generate it and then later on import that transformed variable into HLM:

```
. gen sizesqrt=sqrt(size)
```

If a variable contains zero or negative values, you need to add a constant to it before looking for transformations (such that all values of the variable become larger than zero). For example:

```
. ladder mathach

Transformation         Formula              chi2(2)  Prob > chi2
-----------------------------------------------------------------
Cubic                  mathach^3             758.08        0.000
Square                 mathach^2             758.90        0.000
Identity               mathach               914.19        0.000
Square root            sqrt(mathach)              .            .
Log                    log(mathach)               .            .
1/(Square root)        1/sqrt(mathach)            .            .
Inverse                1/mathach                  .            .
1/Square               1/(mathach^2)              .            .
1/Cubic                1/(mathach^3)              .            .

. gladder mathach

. sum mathach

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
     mathach |      7,185    12.74785    6.878246     -2.832     24.993

. gen mathach_c=mathach-r(min)+1

. sum mathach_c

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   mathach_c |      7,185    16.57985    6.878246          1     28.825

. ladder mathach_c

Transformation         Formula              chi2(2)  Prob > chi2
-----------------------------------------------------------------
Cubic                  mathac~c^3            595.26        0.000
Square                 mathac~c^2           1029.42        0.000
Identity               mathac~c              914.19        0.000
Square root            sqrt(mathac~c)        380.40        0.000
Log                    log(mathac~c)              .            .
1/(Square root)        1/sqrt(mathac~c)           .            .
Inverse                1/mathac~c                 .            .
1/Square               1/(mathac~c^2)             .            .
1/Cubic                1/(mathac~c^3)             .            .

. gladder mathach_c
```

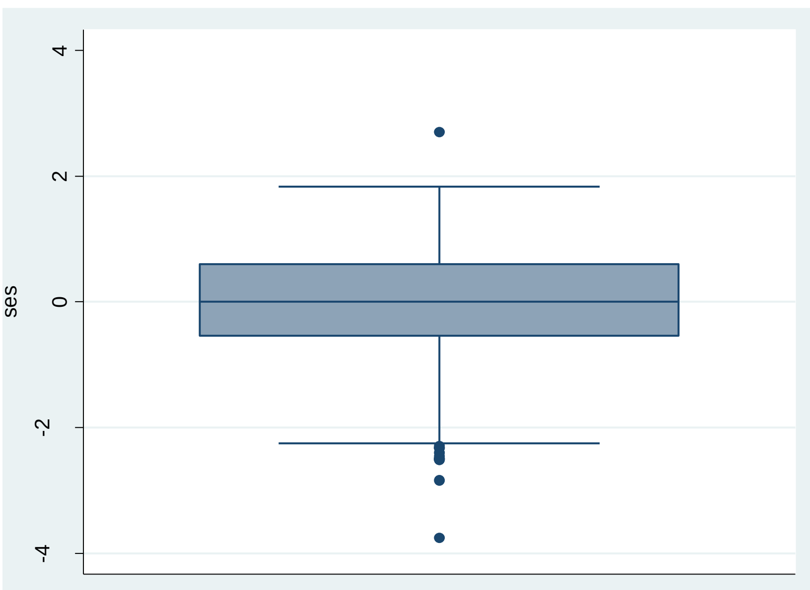
If your sample size is large, everything will be significantly different from normal, so you should either rely on graphical examination (gladder) or randomly select a subsample of your dataset and do this type of analysis for that subsample.

If as variable is negatively skewed, you might have an easier time finding a transformation for it after reversing it. To reverse the variable and yet keep all the values positive, you can subtract it from its maximum value +1; for example:

```
. sum mathach

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     mathach |      7185    12.74785    6.878246     -2.832     24.993

. gen mathachr=r(max)+1-mathach

. sum mathachr

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    mathachr |      7185    13.24515    6.878246   .9999999     28.825
```

As you are examining normality, pay attention to outliers as well – sometimes, it is useful to top-code or bottom-code outliers in addition to or instead of transforming a variable.

```
. graph box ses
```




```
. sum ses, detail
                             ses
-------------------------------------------------------------
      Percentiles      Smallest
 1%       -1.848         -3.758
 5%       -1.318         -2.838
10%       -1.038         -2.508       Obs                7185
25%        -.538         -2.508       Sum of Wgt.        7185

50%         .002                      Mean           .0001434
                        Largest       Std. Dev.      .7793552
75%         .602          1.732
90%        1.022          1.762       Variance       .6073945
95%        1.212          1.832       Skewness      -.2281447
99%        1.512          2.692       Kurtosis       2.620279
```

```
. gen ses1=clip(ses, -2.9, 1.9)
```

Never top-code or bottom-code more than 5% of the distribution; better yet, do 1% or less. Sometimes transformation might be a better way to bring in outliers so consider both options or a combination of them.

If you do a good job dealing with normality problems and with outliers during preliminary screening, you should not run into problems with multivariate normality. Still, we need to check both level 1 and level 2 residuals for normality. Let's estimate a model, obtain residuals, and inspect them.

```
. mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses
i.female##i.sector || id: female, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(8)      =     849.53
Log likelihood = -23218.854                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.856887   .1473012    19.39   0.000     2.568182    3.145592
     meanses |   3.203604   .4968255     6.45   0.000     2.229844    4.177364
             |
       c.ses#|
   c.meanses |    .832274    .268213     3.10   0.002     .3065862    1.357962
             |
         ses |          0  (omitted)
    1.sector |   1.167355   .3994507     2.92   0.003     .3844458    1.950264
             |
sector#c.ses |
          1  |  -1.554133   .2223377    -6.99   0.000    -1.989907   -1.118359
             |
    1.female |   -1.22104    .238342    -5.12   0.000    -1.688182   -.7538981
     meanses |          0  (omitted)
             |
      female#|
   c.meanses |
          1  |  -.0074533   .5063542    -0.01   0.988    -.9998894    .9849827
             |
      female#|
      sector |
        1 1  |    .047171   .4186158     0.11   0.910     -.773301    .8676429
             |
       _cons |   12.71993   .2439351    52.14   0.000     12.24183    13.19804
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                 var(female) |   .9446425   .5677046       .290883    3.067726
                  var(_cons) |   2.897432     .55434      1.991409    4.215665
           cov(female,_cons) |  -1.091921   .4762149     -2.025285   -.1585568
```

```
-----------------------------+------------------------------------------------
               var(Residual) |   36.37821   .6188368      35.18531    37.61155
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 190.77                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

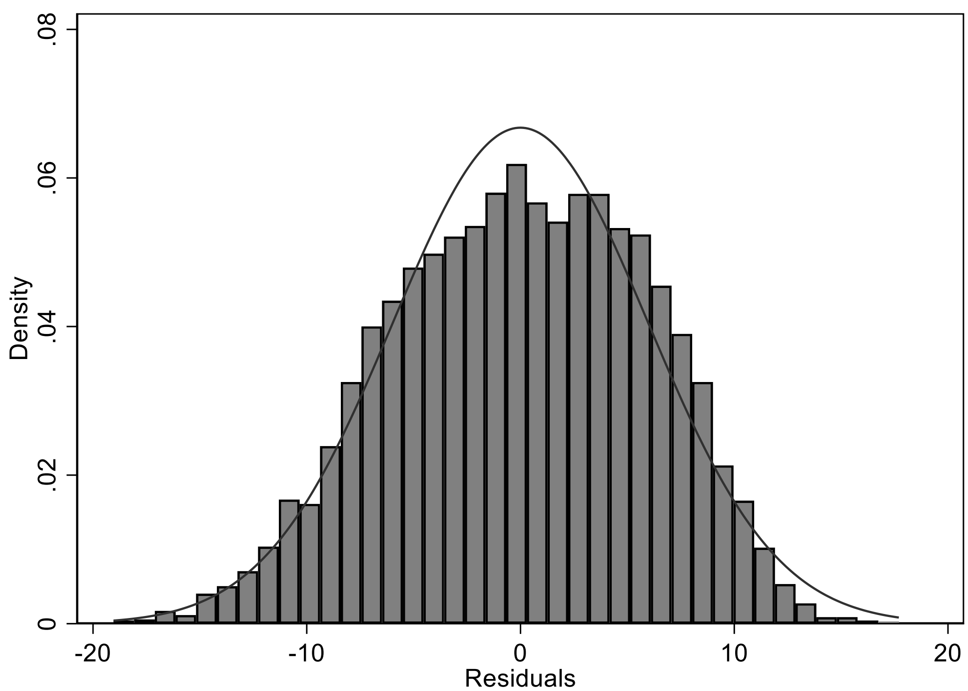
To check the distribution of error terms, we obtain level-1 residuals and level-2 residuals. Here is what we can obtain using predict command:

```
      xb           xb, linear predictor for the fixed portion of the model
      stdp         standard error of the fixed-portion linear prediction xb
      fitted       fitted values, linear predictor of the fixed portion plus
                     contributions based on predicted random effects
      residuals    residuals, response minus fitted values
      rstandard    standardized residuals
      reffects     best linear unbiased predictions (BLUPs) of the
                   random effects. By default, BLUPs for all random effects in the
                   model are calculated. You must specify q new variables, where q is
                   the number of random-effects terms in the model.
      reses         standard errors of the best linear unbiased predictions (BLUPs) of
                   the random effects.  By default, standard errors for all BLUPs in
                    the model are calculated. You must specify q new variables.
```

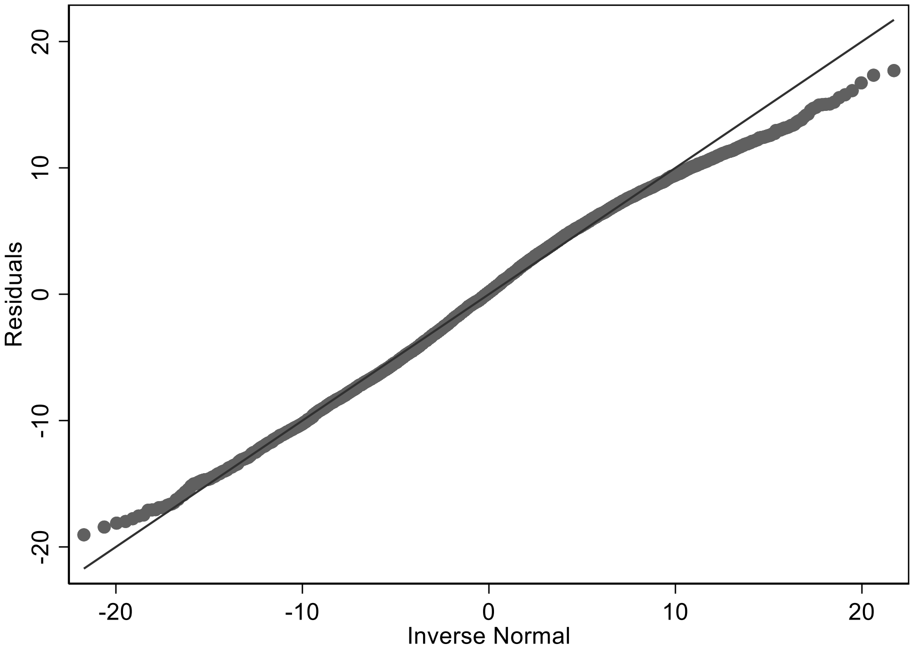
Thus, residuals will give you level 1 residuals, and reffects will give you level 2 residuals for each level 2 random component. You should examine both types of residuals to assess normality.

Level-1 residuals:

```
. predict l1resid, resid

. histogram l1resid, normal
(bin=38, start=-19.084782, width=.96868813)
```

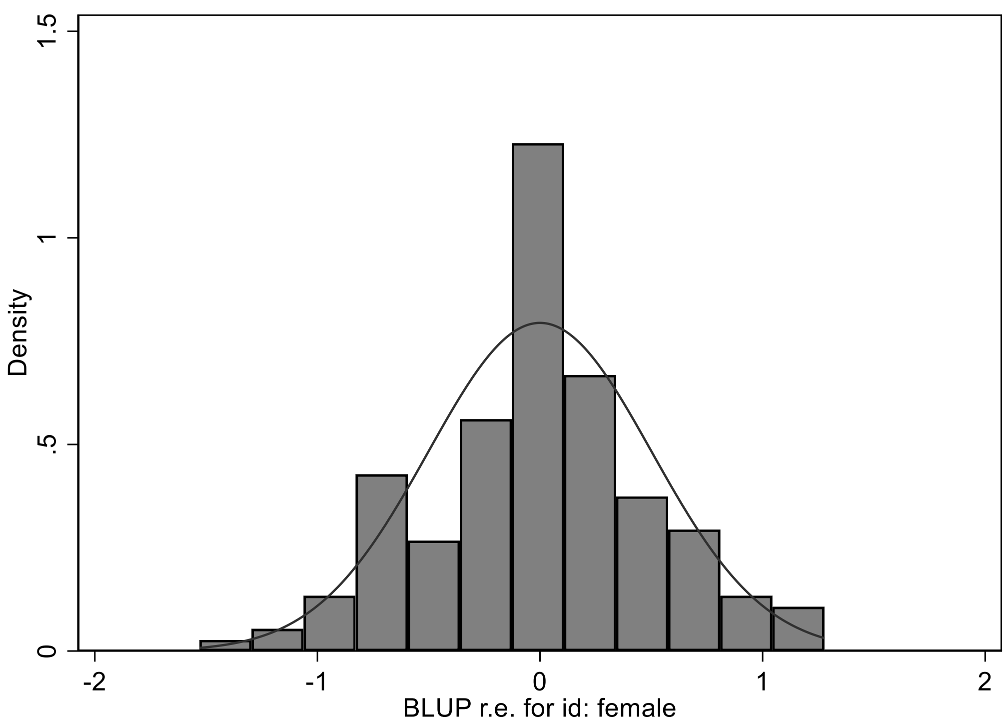


```
. qnorm l1resid
```

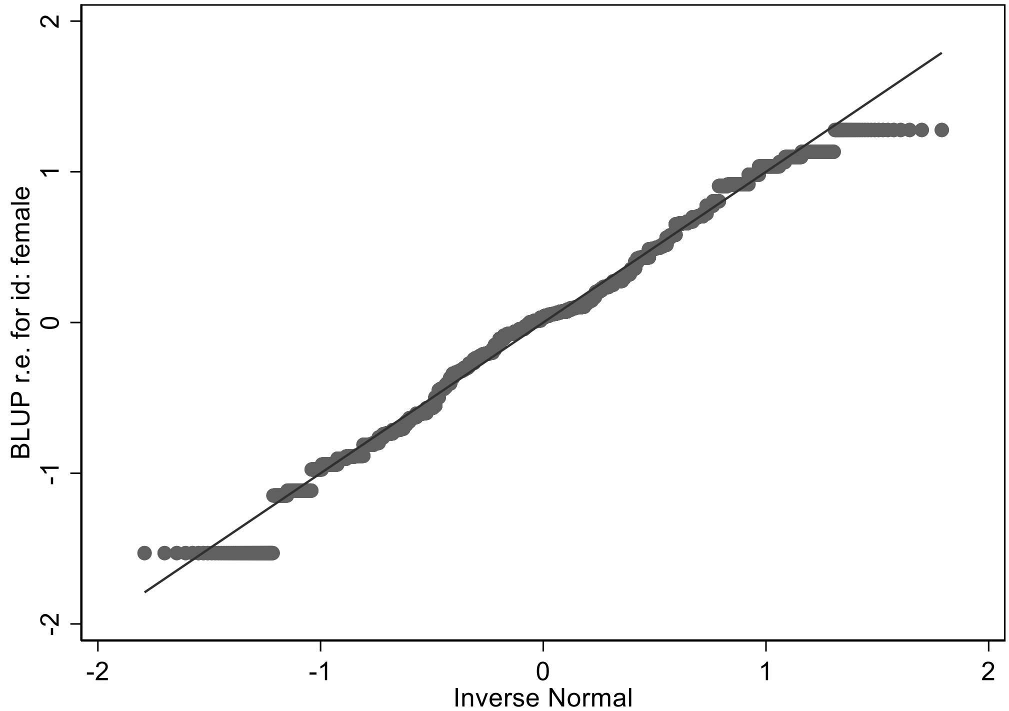


Level 2 residuals:

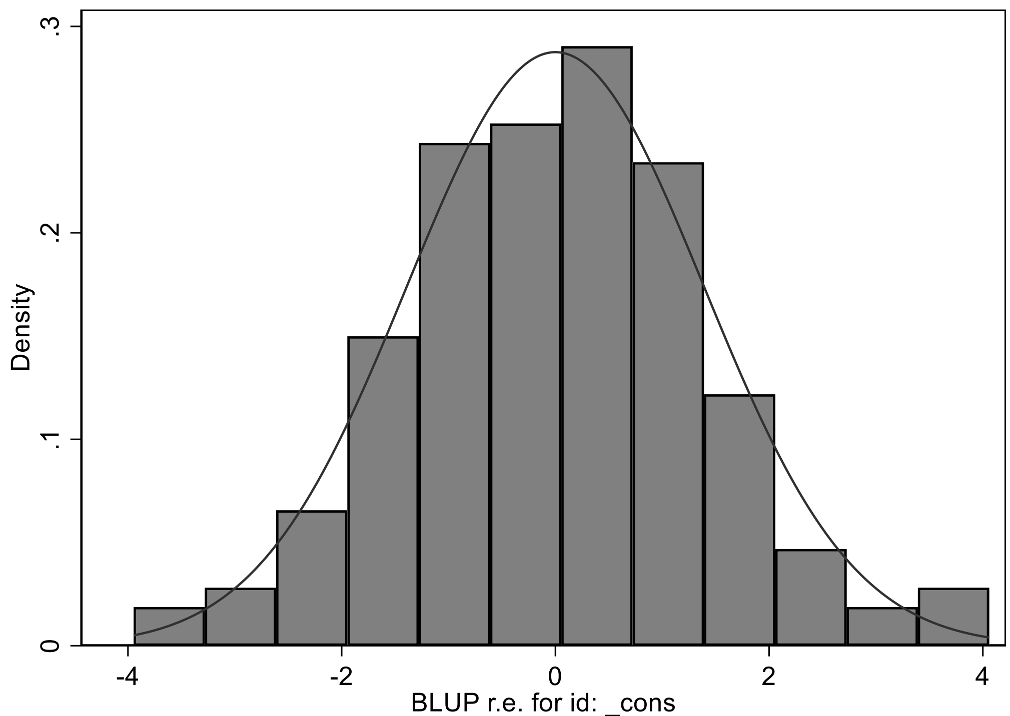
```
. predict l2resid*, reffects

. histogram  l2resid1 if tagged==1, normal
(bin=12, start=-4.0914528, width=.6652911)
```



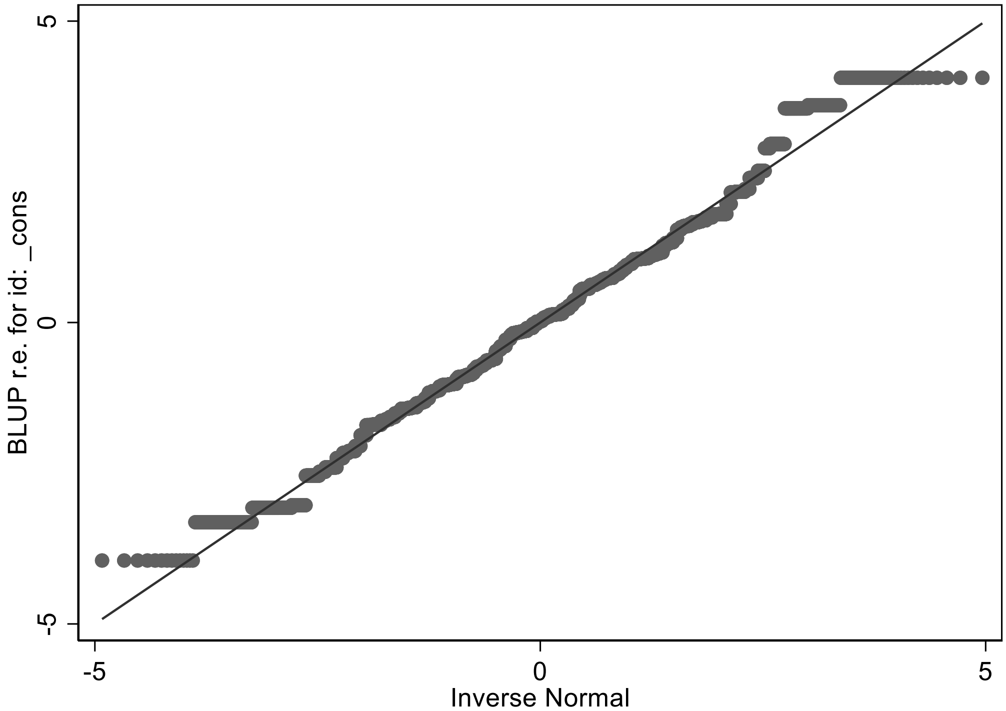

```
. qnorm l2resid1
```




```
. histogram   l2resid2 if tagged==1, normal
(bin=12, start=-1.3546074, width=.2191729)
```




```
. qnorm l2resid2
```



If there are problems with normality of residuals but you can't fix them with simple transformations or top/bottomcoding, you can use robust option available with mixed for robust standard errors to obtain standard errors and significance tests that are less dependent on assumptions:

```
. mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses i.female#
> #i.sector || id: female, cov(unstr) robust
note: ses omitted because of collinearity
note: meanses omitted because of collinearity

Mixed-effects regression                        Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(8)      =    1204.55
Log pseudolikelihood = -23218.854               Prob > chi2       =     0.0000

                                   (Std. Err. adjusted for 160 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.856887   .1408367    20.29   0.000     2.580852    3.132922
     meanses |   3.203604   .4716531     6.79   0.000      2.27918    4.128027
             |
      c.ses#|
   c.meanses |    .832274   .2933011     2.84   0.005     .2574144    1.407134
             |
         ses |          0  (omitted)
```

```
    1.sector |   1.167355   .3975957     2.94   0.003     .3880816    1.946628
             |
sector#c.ses |
          1  |  -1.554133   .2228935    -6.97   0.000    -1.990996    -1.11727
             |
    1.female |   -1.22104   .2216856    -5.51   0.000    -1.655536   -.7865441
     meanses |          0  (omitted)
             |
      female#|
   c.meanses |
          1  |  -.0074533   .4225481    -0.02   0.986    -.8356324    .8207257
             |
      female#|
      sector |
        1 1  |    .047171   .4142162     0.11   0.909    -.7646779    .8590199
             |
       _cons |   12.71993   .2224148    57.19   0.000     12.28401    13.15586
------------------------------------------------------------------------------
------------------------------------------------------------------------------
                             |                 Robust
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                 var(female) |   .9446425   .5153373      .3242692    2.751879
                  var(_cons) |   2.897432   .5871549      1.947691    4.310289
           cov(female,_cons) |  -1.091921    .452792     -1.979377   -.2044647
-----------------------------+------------------------------------------------
               var(Residual) |   36.37821   .7087571      35.01526    37.79421
------------------------------------------------------------------------------
```

You can also use bootstrapping, although it does take time to calculate:

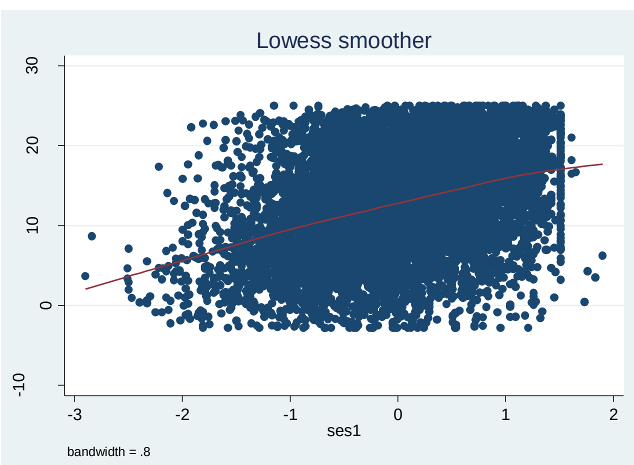
```
. bootstrap, cluster(id): mixed mathach c.ses##c.meanses c.ses##i.sector
i.female##c.meanses i.female##i.sector || id: female, cov(unstr)
(running mixed on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(8)      =    1548.18
Log likelihood = -23218.854                     Prob > chi2       =     0.0000

                                    (Replications based on 160 clusters in id)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.856887   .1349074    21.18   0.000     2.592474    3.121301
     meanses |   3.203604   .3618389     8.85   0.000     2.494412    3.912795
             |
      c.ses#|
   c.meanses |    .832274   .3020349     2.76   0.006     .2402965    1.424252
             |
         ses |          0  (omitted)
```

```
    1.sector |   1.167355   .3569849     3.27   0.001     .4676774    1.867032
             |
sector#c.ses |
          1  |  -1.554133   .2129321    -7.30   0.000    -1.971472   -1.136794
             |
    1.female |   -1.22104   .1914944    -6.38   0.000    -1.596362   -.8457178
    meanses  |          0  (omitted)
             |
     female#|
   c.meanses |
          1  |  -.0074533   .3536247    -0.02   0.983     -.700545    .6856383
             |
     female#|
     sector |
        1 1  |    .047171   .3585003     0.13   0.895    -.6554766    .7498186
             |
       _cons |   12.71993   .1796728    70.79   0.000     12.36778    13.07209
------------------------------------------------------------------------------

------------------------------------------------------------------------------
                             |   Observed   Bootstrap         Normal-based
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                 var(female) |   .9446425   .2268802      .5899694    1.512535
                  var(_cons) |   2.897432   .4663447      2.113545    3.972053
           cov(female,_cons) |  -1.091921   .2330362     -1.548663   -.6351782
-----------------------------+------------------------------------------------
               var(Residual) |   36.37821   .6943167      35.04251    37.76482
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 190.77                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

4. Linearity

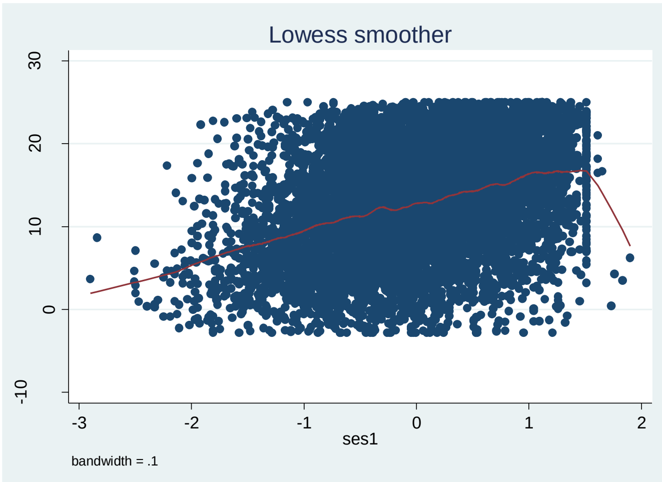
Before running mixed models, it's also a good idea to examine the relationship of each independent variable to the dependent to assess its linearity. A good tool for such an examination is a lowess plot – that is, a scatterplot with locally weighted regression line (based on means or medians) going through it:

```
. lowess mathach ses1
```



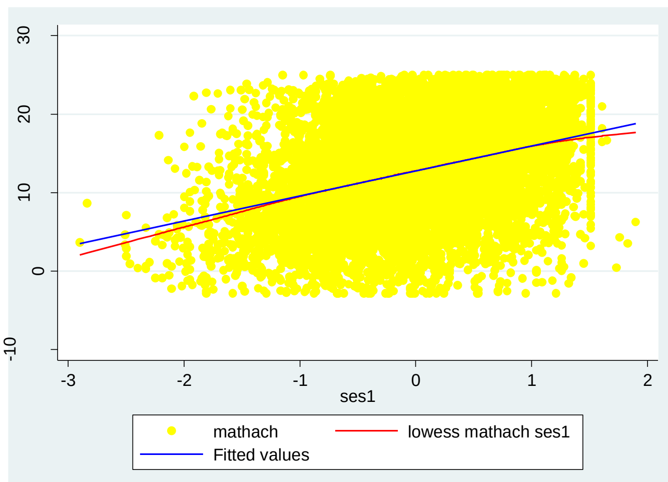
We can change bandwidth to make the curve less smooth (decrease the number) or smoother (increase the number):

```
. lowess mathach ses1, bwidth(.1)
```



We can also add a regression line to see the difference better:

```
. scatter mathach ses1, mcolor(yellow) || lowess mathach ses1, lcolor(red) ||
lfit mathach ses1, lcolor(blue)
```

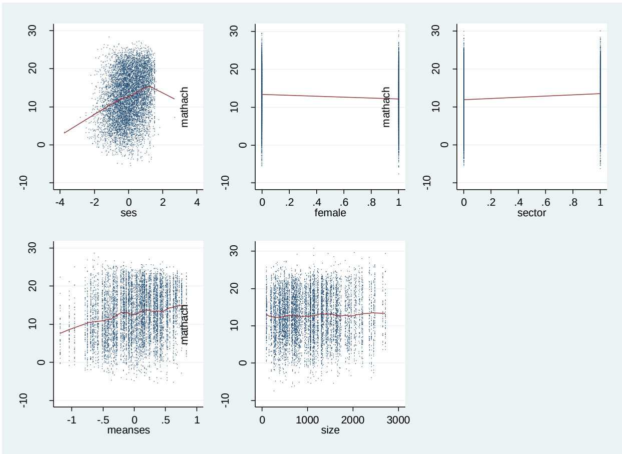


You can do an approximate test for multivariate linearity (based on OLS) with a user-written mrunning program:

```
. search mrunning
Keyword search
        Keywords:  mrunning
          Search:  (1) Official help files, FAQs, Examples, SJs, and STBs
Search of official help files, FAQs, Examples, SJs, and STBs
SJ-5-3  gr0017  . . . . . . . . . . . . . . A multivariable scatterplot smoother
        (help mrunning, running if installed) . . . . P. Royston and N. J. Cox
        Q3/05   SJ 5(3):405--412
        presents an extension to running for use in a
        multivariable context
```

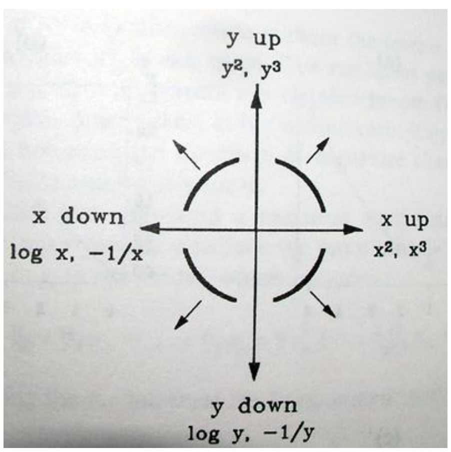
Click on gr0017 to install the program. Now we can use it:

```
. mrunning mathach ses female sector meanses size
```



If the relationship looks nonlinear on lowess plot, consider using transformations to fix it before importing data into HLM. (Note that if the relationship is too complex, sometimes we may choose to break up the corresponding independent variable into a series of dummies instead.)

Monotone nonlinear relationship: Power transformations can be used to linearize relationships if strong monotone nonlinearities are found. The following chart gives suggestions for transformations when the curve looks a certain way:



Nonmonotone relationship: For non-monotone relationships (e.g. parabola or cubic), use polynomial functions of the variable, e.g. ses and ses squared, etc. Note that when including variables that are generated using other variables already in the model (as in this case, or when we want to enter a product of two variables to model an interaction term), we should mean-center the variable outside of HLM (only if it is continuous; don't mean-center dichotomous variables!), and then square and/or cube the mean-centered variable. We will then include the mean-centered variable itself and its transformations into our HLM file and our models. For example, if we are dealing with a second level variable, we would get its mean across 160 level 2 cases by restricting the calculation to one case per level 2 unit:

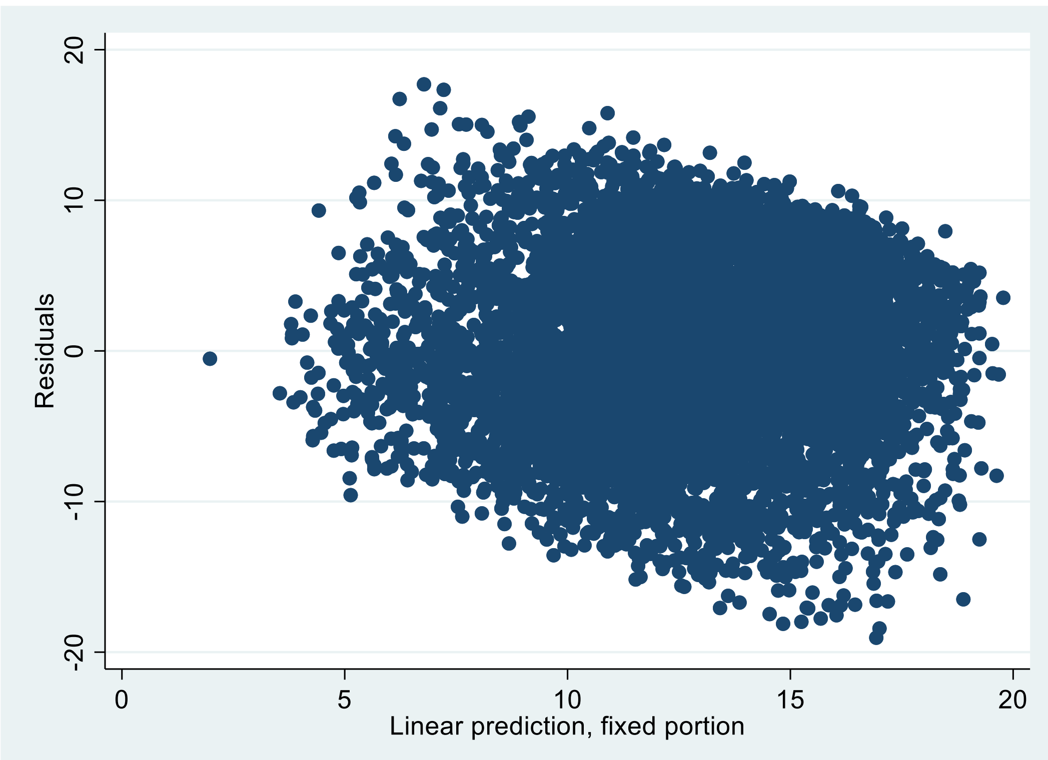
```
. sum size if tagged==1
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
```

```
      size |        160    1097.825    629.5064        100       2713

. gen sizem=size-r(mean)

. gen sizem2=sizem^2
```

Oftentimes, the same transformation that helps with normality also will improve linearity, but that it is not always the case. Overall, linearity is more important to enforce than normality for a given variable, so if you end up with incompatible transformations, opt for the one improving linearity.

Once we estimated our HLM model and obtained residuals, we can inspect them to further assess linearity. First, we can assess the overall pattern by plotting level 1 residuals against predicted values; there should be no discernable pattern:

```
. qui mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses
i.female##i.sector || id: female, cov(unstr)

. predict xb, xb

. scatter l1resid xb
```

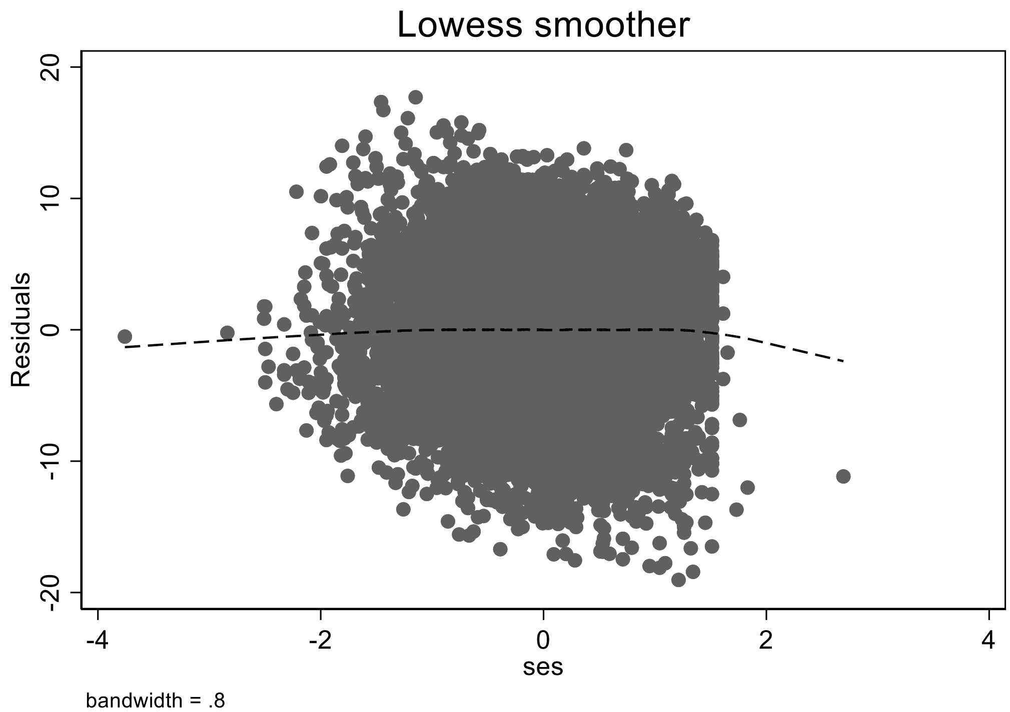


This does not look too good; indicates potential heteroscedasticity or nonlinearity problems.

To test the linearity assumption for continuous predictors, it is useful to plot residuals against each of the continuous dependent variable. To improve our ability to detect a curvilinear relationship, we will include a smoother in our plot using lowess command.

In level 1 file:

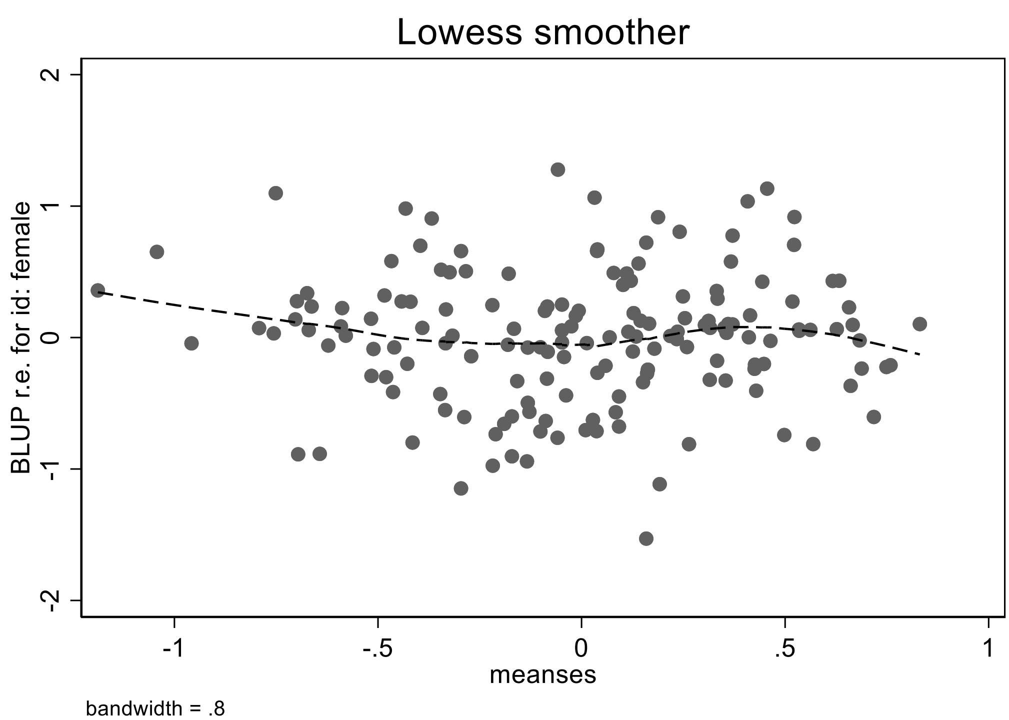
```
. lowess l1resid ses
```



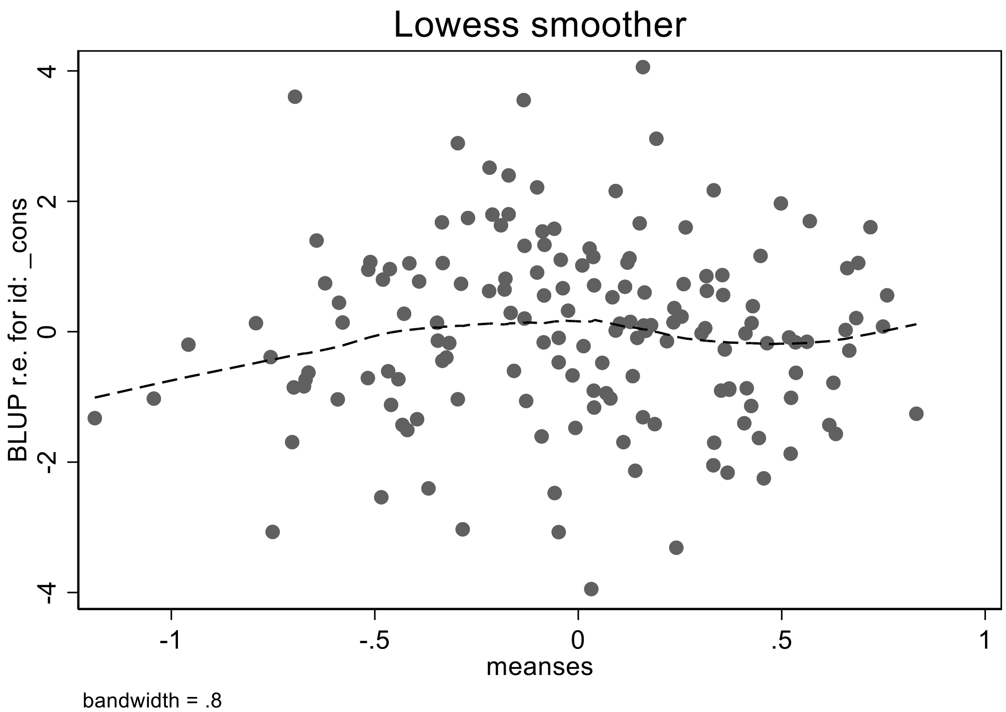
Looks more or less fine, but we do see some outliers on SES.

In level 2 file:

```
. lowess l2resid1 meanses if tagged==1
```



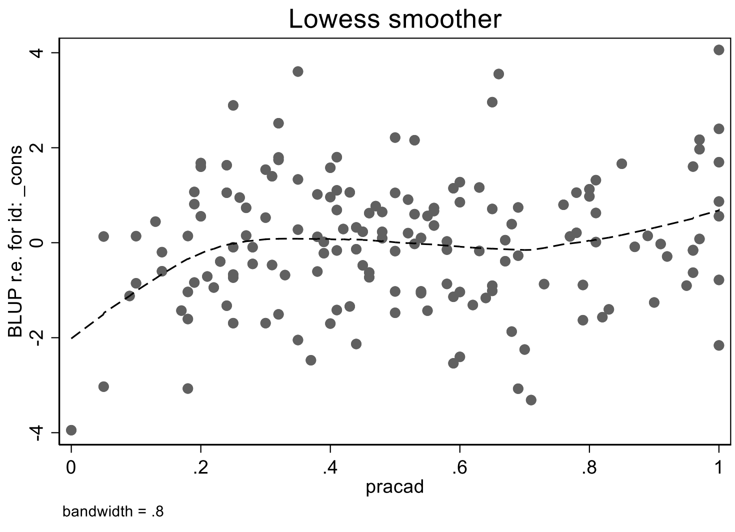

```
. lowess l2resid2 meanses if tagged==1
```



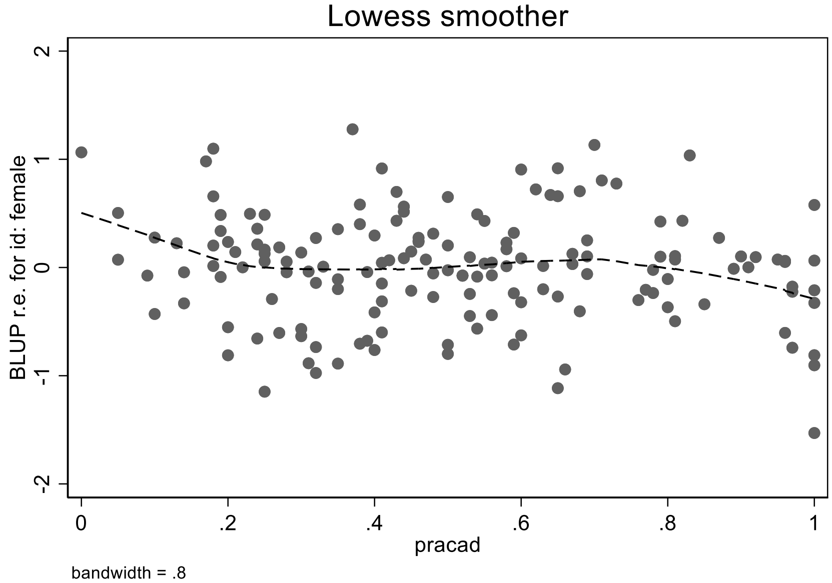
Based on these graphs, we could consider modeling nonlinear relationships with MEANSES (e.g. cubic).

We can also use such plots to search for potential other relationships and examine their shape, e.g. with PRACAD:

```
. lowess l2resid1 pracad if tagged==1
```




```
. lowess l2resid2 pracad if tagged==1
```



Based on graphs for MEANSES, looks like we need a quadratic and cubic term; let's try that. We run the following model:

Level-1 Model

$MATHACH_{ij} = \beta_{0j} + \beta_{1j}*(FEMALE_{ij}) + \beta_{2j}*(SES_{ij}) + r_{ij}$

Level-2 Model

$\beta_{0j} = \gamma_{00} + \gamma_{01}*(SECTOR_j) + \gamma_{02}*(MEANSESM_j) + \gamma_{03}*(MEANSES2_j) + \gamma_{04}*(MEANSES3_j) + u_{0j}$
$\beta_{1j} = \gamma_{10} + \gamma_{11}*(SECTOR_j) + \gamma_{12}*(MEANSESM_j) + \gamma_{13}*(MEANSES2_j) + \gamma_{14}*(MEANSES3_j) + u_{1j}$
$\beta_{2j} = \gamma_{20} + \gamma_{21}*(SECTOR_j) + \gamma_{22}*(MEANSESM_j) + \gamma_{23}*(MEANSES2_j) + \gamma_{24}*(MEANSES3_j) + u_{2j}$

Mixed Model

$MATHACH_{ij} = \gamma_{00} + \gamma_{01}*SECTOR_j + \gamma_{02}*MEANSESM_j + \gamma_{03}*MEANSES2_j$
$+ \gamma_{04}*MEANSES3_j$
$+ \gamma_{10}*FEMALE_{ij} + \gamma_{11}*SECTOR_j*FEMALE_{ij} + \gamma_{12}*MEANSESM_j*FEMALE_{ij} + \gamma_{13}*MEANSES2_j*FEMALE_{ij}$
$+ \gamma_{14}*MEANSES3_j*FEMALE_{ij}$
$+ \gamma_{20}*SES_{ij} + \gamma_{21}*SECTOR_j*SES_{ij} + \gamma_{22}*MEANSESM_j*SES_{ij} + \gamma_{23}*MEANSES2_j*SES_{ij}$
$+ \gamma_{24}*MEANSES3_j*SES_{ij}$
$+ u_{0j} + u_{1j}*FEMALE_{ij} + u_{2j}*SES_{ij} + r_{ij}$

```
. mixed mathach c.ses##c.meanses##c.meanses##c.meanses c.ses##i.sector
i.female##c.meanses##c.meanses##c.meanses i.female##i.sector || id: female, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
```

```
                                                              max =         67

                                                Wald chi2(14)     =     876.38
Log likelihood = -23211.143                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   3.050629   .1656633    18.41   0.000     2.725935    3.375323
     meanses |   1.817196   .8053935     2.26   0.024     .2386537    3.395738
             |
      c.ses#|
   c.meanses |   .4610032   .4533027     1.02   0.309    -.4274538     1.34946
             |
  c.meanses#|
   c.meanses |  -.9934494   1.086095    -0.91   0.360    -3.122156    1.135258
             |
      c.ses#|
  c.meanses#|
   c.meanses |   -1.44887   .6373161    -2.27   0.023    -2.697987   -.1997537
             |
  c.meanses#|
  c.meanses#|
   c.meanses |   3.866993   1.659089     2.33   0.020     .6152377    7.118749
             |
      c.ses#|
  c.meanses#|
  c.meanses#|
   c.meanses |   .8161978   .9309449     0.88   0.381    -1.008421    2.640816
             |
         ses |          0  (omitted)
    1.sector |   1.283584   .3963461     3.24   0.001     .5067598    2.060408
             |
sector#c.ses |
          1  |    -1.4467   .2256309    -6.41   0.000    -1.888929   -1.004472
             |
    1.female |  -1.230308   .2740098    -4.49   0.000    -1.767358    -.693259
     meanses |          0  (omitted)
             |
     female#|
   c.meanses |
          1  |   .7925874   .8138469     0.97   0.330    -.8025233    2.387698
             |
     female#|
  c.meanses#|
   c.meanses |
          1  |  -.1274134   1.087919    -0.12   0.907    -2.259695    2.004869
             |
     female#|
  c.meanses#|
  c.meanses#|
   c.meanses |
          1  |  -1.805483   1.583418    -1.14   0.254    -4.908926    1.297959
             |
     female#|
      sector |
        1 1  |  -.0036908   .4203907    -0.01   0.993    -.8276415    .8202598
             |
       _cons |    12.8906   .2731983    47.18   0.000     12.35514    13.42606
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
```

```
                var(female) |     .91266   .5645054      .2715297    3.067614
                 var(_cons) |   2.727112    .530815      1.862182    3.993778
          cov(female,_cons) |  -1.011546   .4651488     -1.923221   -.0998711
-----------------------------+------------------------------------------------
              var(Residual) |   36.32734   .6179661      35.13612    37.55895
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 180.35                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

It looks like there is a bunch of non-significant coefficients that we could omit; let's omit and compare using BIC (we could also do joint hypothesis test for these coefficients first):

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .  -23211.14      19   46460.29      46591
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. mixed mathach c.ses##c.meanses##c.meanses##c.meanses c.ses##i.sector i.female || id:
female, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(10)     =     873.56
Log likelihood = -23212.015                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   3.053147   .1655297    18.44   0.000     2.728715    3.377579
     meanses |   2.324254   .6047559     3.84   0.000     1.138954    3.509553
             |
       c.ses#|
   c.meanses |   .4447929   .4528177     0.98   0.326    -.4427134    1.332299
             |
   c.meanses#|
   c.meanses |  -1.088537   .8269356    -1.32   0.188    -2.709301    .5322275
             |
       c.ses#|
   c.meanses#|
   c.meanses |  -1.458181   .6368045    -2.29   0.022    -2.706295   -.2100674
             |
   c.meanses#|
   c.meanses#|
   c.meanses |   2.697434   1.288303     2.09   0.036     .1724068    5.222461
             |
       c.ses#|
   c.meanses#|
   c.meanses#|
   c.meanses |   .8568271   .9304757     0.92   0.357    -.9668717    2.680526
             |
```

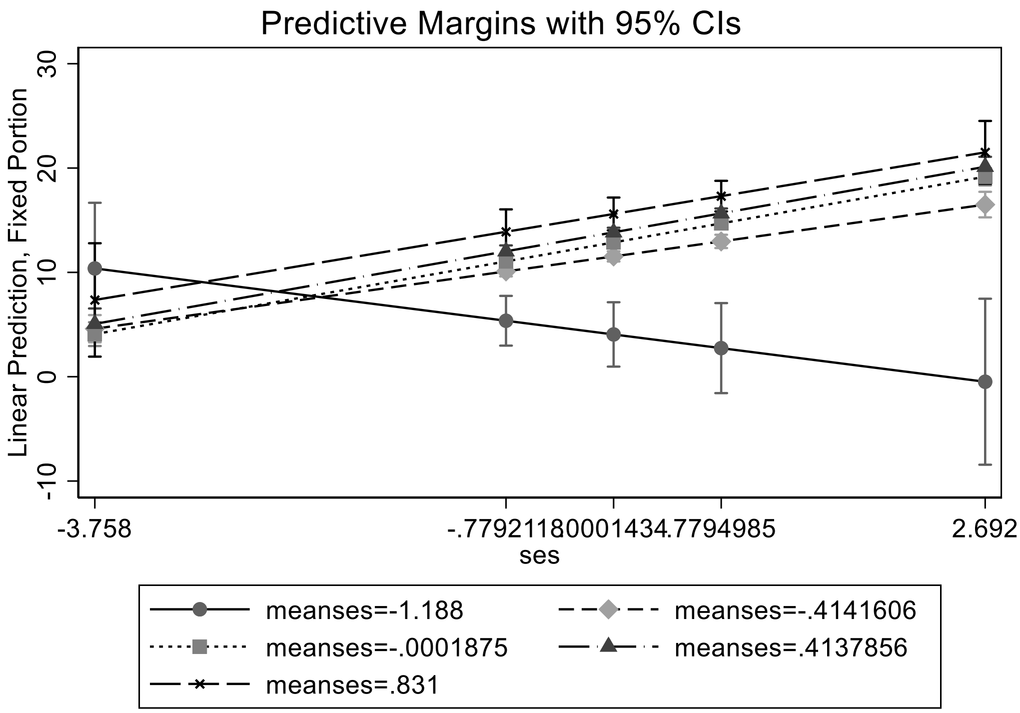
```
         ses |          0  (omitted)
    1.sector |   1.285378   .2859008     4.50   0.000     .7250229    1.845734
             |
sector#c.ses |
          1  |  -1.447254   .2254075    -6.42   0.000    -1.889045   -1.005464
             |
    1.female |  -1.213093   .1813475    -6.69   0.000    -1.568528   -.8576586
       _cons |   12.87913   .2388608    53.92   0.000     12.41097    13.34729
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                 var(female) |   .9800848   .5711366      .3127755    3.071104
                  var(_cons) |   2.742527   .5344897      1.871805    4.018289
           cov(female,_cons) |  -1.047902   .4702886     -1.969651   -.1261537
-----------------------------+------------------------------------------------
               var(Residual) |    36.3267   .6179416      35.13552    37.55826
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 182.23                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .  -23212.02      15   46454.03   46557.23
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

BIC strongly prefers the more parsimonious model. We could also consider getting rid of cubed term for MEANSES, but let's keep for now and examine graphically using margins and marginsplot:

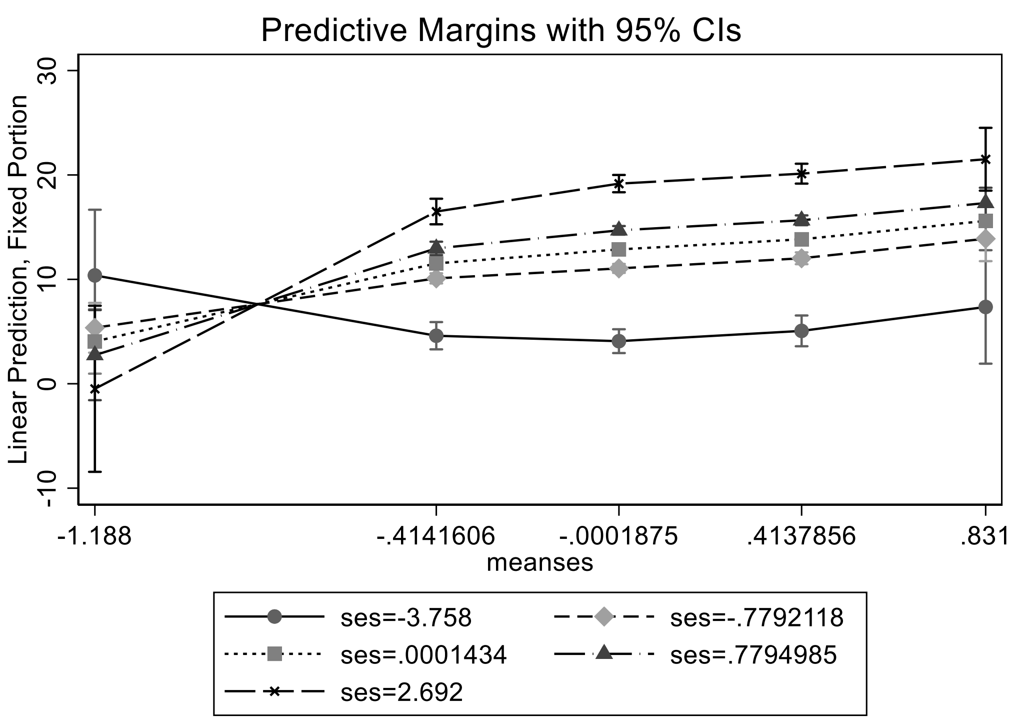
```
. sum ses

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         ses |      7,185    .0001434    .7793552     -3.758      2.692

. global sesmin=r(min)

. global sesmax=r(max)

. global sesmean=r(mean)

. global plussd=r(mean)+r(sd)

. global minussd=r(mean)-r(sd)

. sum meanses if tagged==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     meanses |        160   -.0001875    .4139731     -1.188       .831

. global meansesmin=r(min)
```

```
. global meansesmax=r(max)

. global meansesmean=r(mean)

. global meansesplussd=r(mean)+r(sd)

. global meansesminussd=r(mean)-r(sd)

. margins, at(ses= ($sesmin $minussd $sesmean $plussd $sesmax) meanses=($means
> esmin $meansesminussd $meansesmean $meansesplussd $meansesmax))

Predictive margins                              Number of obs     =      7,185

Expression   : Linear prediction, fixed portion, predict()

1._at        : ses             =      -3.758
               meanses         =      -1.188

2._at        : ses             =      -3.758
               meanses         =   -.4141606

3._at        : ses             =      -3.758
               meanses         =   -.0001875

4._at        : ses             =      -3.758
               meanses         =    .4137856

5._at        : ses             =      -3.758
               meanses         =        .831

6._at        : ses             =   -.7792118
               meanses         =      -1.188

7._at        : ses             =   -.7792118
               meanses         =   -.4141606

8._at        : ses             =   -.7792118
               meanses         =   -.0001875

9._at        : ses             =   -.7792118
               meanses         =    .4137856

10._at       : ses             =   -.7792118
               meanses         =        .831

11._at       : ses             =    .0001434
               meanses         =      -1.188

12._at       : ses             =    .0001434
               meanses         =   -.4141606

13._at       : ses             =    .0001434
               meanses         =   -.0001875

14._at       : ses             =    .0001434
               meanses         =    .4137856

15._at       : ses             =    .0001434
               meanses         =        .831

16._at       : ses             =    .7794985
               meanses         =      -1.188
```

```
17._at       : ses            =    .7794985
               meanses        =   -.4141606

18._at       : ses            =    .7794985
               meanses        =   -.0001875

19._at       : ses            =    .7794985
               meanses        =    .4137856

20._at       : ses            =    .7794985
               meanses        =        .831

21._at       : ses            =       2.692
               meanses        =      -1.188

22._at       : ses            =       2.692
               meanses        =   -.4141606

23._at       : ses            =       2.692
               meanses        =   -.0001875

24._at       : ses            =       2.692
               meanses        =    .4137856

25._at       : ses            =       2.692
               meanses        =        .831

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   10.37874   3.209518     3.23   0.001       4.0882    16.66928
          2  |   4.600451   .6663532     6.90   0.000     3.294423     5.90648
          3  |   4.080304   .5811327     7.02   0.000     2.941305    5.219303
          4  |   5.065367    .751264     6.74   0.000     3.592916    6.537817
          5  |   7.355452   2.772092     2.65   0.008     1.922252    12.78865
          6  |   5.363829   1.216027     4.41   0.000     2.980459    7.747198
          7  |   10.09419   .2479459    40.71   0.000      9.60822    10.58015
          8  |    11.0489   .2189395    50.47   0.000     10.61979    11.47801
          9  |   12.01957   .2887416    41.63   0.000     11.45365     12.5855
         10  |   13.89045   1.095341    12.68   0.000     11.74362    16.03728
         11  |   4.051753   1.573768     2.57   0.010     .9672248    7.136281
         12  |   11.53154   .2505531    46.02   0.000     11.04046    12.02261
         13  |   12.87213     .18279    70.42   0.000     12.51387    13.23039
         14  |   13.83904   .2253969    61.40   0.000     13.39727    14.28081
         15  |   15.60024    .801719    19.46   0.000      14.0289    17.17158
         16  |   2.739676   2.201374     1.24   0.213    -1.574938    7.054291
         17  |   12.96889   .3257726    39.81   0.000     12.33039    13.60739
         18  |   14.69536   .2105043    69.81   0.000     14.28278    15.10794
         19  |    15.6585   .2385924    65.63   0.000     15.19087    16.12613
         20  |   17.31002   .7472577    23.16   0.000     15.84543    18.77462
         21  |  -.4800974   4.058521    -0.12   0.906    -8.434652    7.474457
         22  |   16.49609    .625075    26.39   0.000     15.27096    17.72121
         23  |   19.16948   .4228441    45.33   0.000     18.34072    19.99824
         24  |   20.12338   .4857143    41.43   0.000      19.1714    21.07536
         25  |   21.50576    1.53443    14.02   0.000     18.49833    24.51318
------------------------------------------------------------------------------

. marginsplot, x(meanses)

  Variables that uniquely identify margins: ses meanses
```




```
. marginsplot, x(ses)

  Variables that uniquely identify margins: ses meanses
```

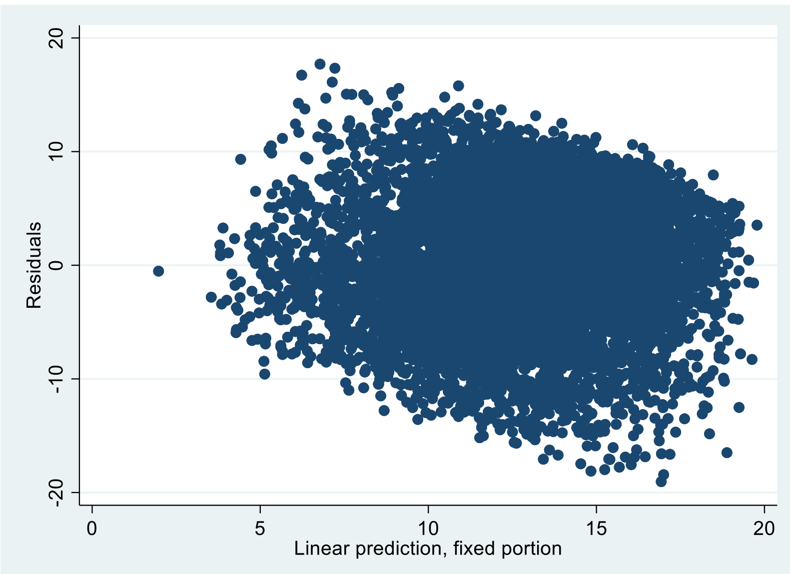


5. Homoscedasticity.

In HLM, the level-1 error terms should have equal variance across level-2 units (the assumption of homoscedasticity or homogeneity of variance) – e.g., all schools should have variances equal to the other schools in the sample.

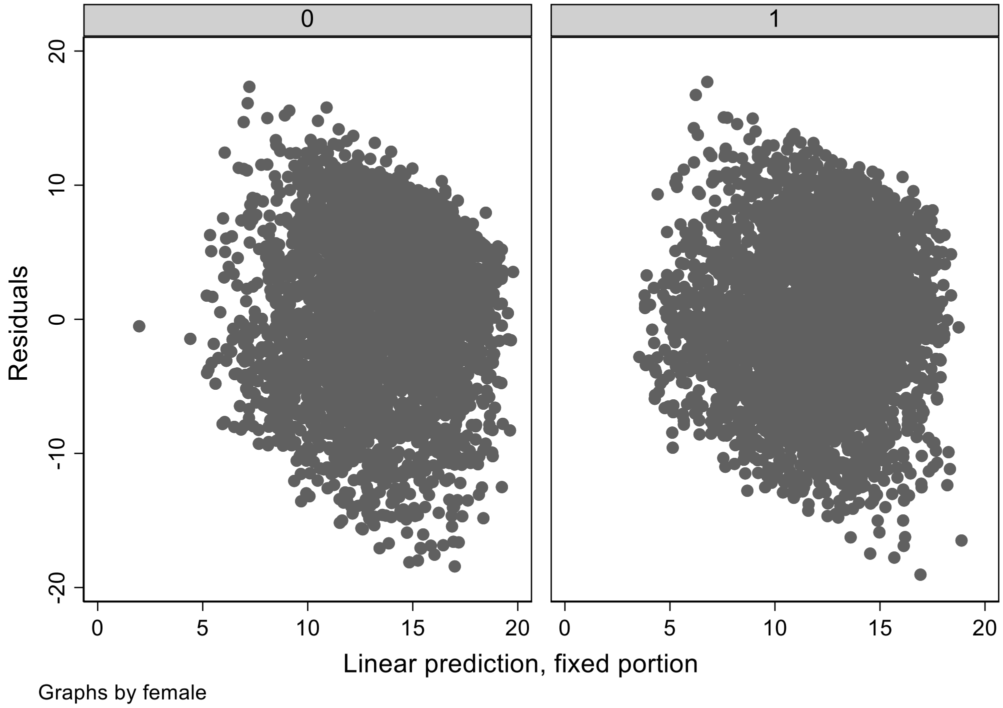
In order to graphically examine whether residual variance is heterogenous, we can look at an RFV plot – residuals vs. fitted values – we already constructed that plot above for the following model:

```
. qui mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses
i.female##i.sector || id: female, cov(unstr)

. scatter l1resid xb
```



We also already started to look at how residuals are distributed along the values of each individual predictor by constructing RVP plots -- residuals vs. predictors – for our continuous variables. However, for dichotomies, such plots are not very informative – however, we could create separate RVF plots by group to assess heterogeneity:

```
. graph twoway (scatter l1resid xb, by(female))
```



We can also assess residual variance by group using sum command:

```
. bysort female: sum l1resid

-------------------------------------------------------------------------------
-> female = 0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     l1resid |      3,390   -1.63e-09    6.137624  -18.42606   17.33541

-------------------------------------------------------------------------------
-> female = 1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     l1resid |      3,795    3.99e-09    5.826746  -19.04129   17.70039
```

Heterogeneity of variance can be a nuisance. When it is a nuisance, the causes can be:

- One or more important level-1 predictors may have been omitted from the model.
- The effects of a level-1 predictor that is random or nonrandomly varying have been erroneously treated as fixed.
- Dependent variable is severely non-normal – skewed or kurtotic (has heavy tails).
- One (or more) of the independent variables has a nonlinear relationship to the dependent variable that we failed to model correctly.
- There are outliers or bad data.

We can deal with each of these problems; if that fails to remove heterogeneity, we can ultimately rely on robust SE. Robust standard errors are standard errors that are relatively insensitive to misspecification at the levels of the model and the distributional assumptions at each level. If the robust and model-based standard errors differ substantially, that suggests that you have some problem with normality, homoscedasticity, or linearity, and you should further investigate those HLM assumptions. If it is not possible to correct the problem, you can report robust standard errors. Note, however, that the robust standard errors should be trusted only when the number of higher-level units is moderately large relative to the number of explanatory variables at the higher level.

Alternatively, heterogeneity of variance can be considered substantively interesting. In that case, we can model it using level 1 predictors – to see whether there are some predictors that seem to explain why level 1 variance is not uniform.

```
. mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses
i.female##i.sector || id: female, cov(unstr) residuals(independent, by(female))
note: ses omitted because of collinearity
note: meanses omitted because of collinearity

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(8)      =     855.62
Log likelihood = -23213.302                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.869241   .1473036    19.48   0.000     2.580531    3.157951
     meanses |   3.193401   .4964206     6.43   0.000     2.220435    4.166368
             |
      c.ses#|
   c.meanses |   .8249061   .2676124     3.08   0.002     .3003954    1.349417
             |
         ses |          0  (omitted)
    1.sector |   1.175409    .398656     2.95   0.003      .394058    1.956761
             |
sector#c.ses |
          1  |  -1.556612   .2219166    -7.01   0.000    -1.991561   -1.121664
             |
    1.female |  -1.219793   .2381591    -5.12   0.000    -1.686577   -.7530101
     meanses |          0  (omitted)
             |
     female#|
   c.meanses |
          1  |  -.0097155   .5062072    -0.02   0.985    -1.001863    .9824323
             |
     female#|
      sector |
        1 1  |   .0393173   .4180634     0.09   0.925    -.7800718    .8587064
             |
       _cons |   12.71989   .2440476    52.12   0.000     12.24157    13.19822
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                 var(female) |   .9168451   .5663447      .2732121    3.076748
```

```
                var(_cons) |    2.79032   .5517962      1.893764     4.11133
         cov(female,_cons) |  -.9889858   .4754416     -1.920834   -.0571373
-----------------------------+------------------------------------------------
Residual: Independent,       |
    by female                |
                 0: var(e) |     38.572   .9571921      36.74083    40.49443
                 1: var(e) |   34.43275   .8047972      32.89096    36.04681
------------------------------------------------------------------------------
LR test vs. linear model: chi2(4) = 201.88                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .   -23213.3      14    46454.6   46550.92
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. est store efemale

. lrtest efemale baseline

Likelihood-ratio test                                 LR chi2(1)  =     11.10
(Assumption: baseline nested in efemale)              Prob > chi2 =    0.0009

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

Looks like gender explains some heterogeneity – there is a higher amount of unexplained variance in math achievement among boys. We could assess model fit:

```
. est store egender

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
     egender |      7,185          .   -23213.3      14    46454.6   46550.92
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. qui mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses
i.female##i.sector || id: female, cov(unstr)

. est store baseline

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
```

```
           . |      7,185          .  -23218.85      13   46463.71   46553.14
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. lrtest baseline egender

Likelihood-ratio test                                 LR chi2(1)  =     11.10
(Assumption: . nested in egender)                     Prob > chi2 =    0.0009

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```

LR test suggests that the model with heterogenous variance is preferred; however, BIC difference is only approximately 3, which is positive but not strong evidence in favor of that more complex model. We might want to explore what explains that higher unexplained variance among boys -- e.g., we could consider an interaction term of SES with gender (which would be an interaction of two level 1 variables), or examine additional predictors such as minority status etc.

We can also allow residual variance to vary by level 2 predictor groups:

```
. mixed mathach c.ses##c.meanses c.ses##i.sector i.female##c.meanses i.female#
> #i.sector || id: female, cov(unstr) residuals(independent, by(sector))
note: ses omitted because of collinearity
note: meanses omitted because of collinearity

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160

                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67

                                                Wald chi2(8)      =     822.03
Log likelihood =  -23205.16                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.855303   .1528454    18.68   0.000     2.555731    3.154874
     meanses |   3.193349   .5003151     6.38   0.000      2.21275    4.173949
             |
       c.ses#|
   c.meanses |   .8257218   .2673987     3.09   0.002       .30163    1.349814
             |
         ses |          0  (omitted)
    1.sector |   1.163307   .3995051     2.91   0.004     .3802913    1.946323
             |
sector#c.ses |
          1  |  -1.550645   .2212156    -7.01   0.000     -1.98422   -1.117071
             |
    1.female |  -1.217677   .2453943    -4.96   0.000    -1.698641   -.7367136
     meanses |          0  (omitted)
             |
      female#|
   c.meanses |
          1  |   .0117788   .5131516     0.02   0.982      -.99398    1.017537
             |
      female#|
```

```
      sector |
        1 1  |    .051641    .417778     0.12   0.902    -.7671888    .8704709
             |
       _cons |    12.7195   .2479454    51.30   0.000     12.23354    13.20547
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                 var(female) |   .9196108    .570399      .2726696    3.101498
                  var(_cons) |   2.901038   .5536637      1.995728    4.217018
           cov(female,_cons) |  -1.080918   .4767802      -2.01539   -.1464463
-----------------------------+------------------------------------------------
Residual: Independent,       |
    by sector                |
                   0: var(e) |   39.55357   .9427131      37.74838    41.44509
                   1: var(e) |   33.13314   .7991106      31.60336    34.73698
------------------------------------------------------------------------------
LR test vs. linear model: chi2(4) = 218.16                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .  -23205.16      14    46438.32   46534.64
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. est store esector

. lrtest baseline esector

Likelihood-ratio test                                 LR chi2(1)  =      27.39
(Assumption: baseline nested in esector)              Prob > chi2 =     0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on
      the boundary of the parameter space.  If this is not true, then the
      reported test is conservative.
```