Annotated WinBUGS Program for LVR-HM3 Model 1

Level-1 Model

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(Grade_{tij} - 7) + \varepsilon_{tij} \qquad \varepsilon_{tij} \sim N(0, \sigma^2) \qquad \text{(Eq.1)}$$

Level-2 Model

$$\pi_{0ij} = \beta_{00j} + r_{0ij} \qquad r_{0ij} \sim N(0, \tau_{\pi 0j}) \qquad \text{(Eq.2a)}$$

$$\pi_{1ij} = \beta_{10j} + \mathrm{Bw}_j(\pi_{0ij} - \beta_{00j}) + r_{1ij} \qquad r_{1ij} \sim N(0, \tau_{\pi 1j}) \qquad \text{(Eq.2b)}$$

$$Cov(r_{0ij}, r_{1ij}) = 0$$

Level-3 Model

$$\beta_{00j} = \gamma_{000} + u_{00j} \qquad u_{00j} \sim N(0, \tau_{\beta 00}) \qquad \text{(Eq.3a)}$$

$$\beta_{10j} = \gamma_{100} + \mathrm{Bb}(\beta_{00j} - \gamma_{000}) + u_{10j} \qquad u_{10j} \sim N(0, \tau_{\beta 10}) \qquad \text{(Eq.3b)}$$

$$\mathrm{Bw}_j = \mathrm{Bw\_0} + \mathrm{Bw\_1}(\beta_{00j} - \gamma_{000}) + u_{\mathrm{Bw}j} \qquad u_{\mathrm{Bw}j} \sim N(0, \tau_{\mathrm{Bw}}) \qquad \text{(Eq.3c)}$$

$$Cov(u_{00j}, u_{10j}) = 0, Cov(u_{00j}, u_{\mathrm{Bw}j}) = 0, Cov(u_{10j}, u_{\mathrm{Bw}j}) = \tau_{\beta 10,\mathrm{Bw}}$$

Note that in WinBUGS, we need to parameterize models in terms of precisions (e.g., $1 / \tau_{\pi 1j}$) rather than variances (e.g., $\tau_{\pi 1j}$). See our JEBS paper as well as the annotated WinBUGS program below for the prior specification for the fixed effects and precisions in the above model .

`WinBUGS Data Format for Model 1`

We begin by inspecting the input files for this analysis. Since our model contains no observed predictors at levels 2 and 3, we will need only two files. The first file (i.e., the level-1 file) contains the time-series observations for the students in our sample. Note that the very first line of this file defines three key constants that are used in our WinBUGS program: T=8585, which is the total number of time-series observations in our sample; I=2628, which is the total number of students in our sample; and J=45, which is the number of schools in our sample.

The vector "Y[]" contains the time series math achievement scores for each of the students in our sample, and "Grade[]" contains the series of grades at which each student was observed, where values of 0, 1, 2 and 3 correspond, respectively, to the start of seventh, eighth, ninth and tenth grade. Both of these vectors are length T=8585.

The level-1 file also contains a third vector of length 8585 named “stu[]”. This vector contains student ID values that link each time-series observation with the appropriate student. Thus as can be seen, the first 4 time-series observations are nested within student 1, the next 4 observations are nested within student 2, and so on. For reasons that will be clear when we examine the code for our programs, the student IDs must range sequentially from 1 to I=2628. As will also be seen, “stu[]” helps us link the time series data for a given student to the initial status and growth rate parameters for that student.

Since there are no observed student-level predictors in this model, the primary role of the level-2 file in this analysis is to link each of the I=2628 students in the sample with the appropriate school that he or she is nested in. Thus “sch[]” is a vector of length 2628 that contains the school ID value for each student. As can be seen, the first 57 students in our sample (students 1-57) are nested in school 1; the next 56 students in our sample (students 58-113) are nested in school 2, and so forth. Note that if there are 45 schools in the sample, then the school IDs must range from 1-45. As will also be seen, “sch[]” plays a key role in specifying the level-2 (within-school) model in our WinBUGS program by linking students in a particular school with, for example, the mean initial status parameter, mean growth rate, and initial status / growth rate slope for that school.

```
## Level-1 File ##
list(T=8585, I = 2628, J=45)
Y[]     stu[]  Grade[]
70.040   1      0
69.510   1      1
71.050   1      2
78.480   1      3
59.360   2      0
65.260   2      1
68.890   2      2
72.040   2      3
68.460   3      0
69.430   3      1
.
.
.
40.190   57     0
40.920   57     1
42.720   58     0
45.250   58     1
.
.
.
53.810   113    0
56.290   113    1
.
.
.
53.630  2528    0
56.600  2528    1
61.460  2528    2
70.150  2528    3
.
.
.
```

```
## Level-2 File ##

sch[]
 1
 1
 1     total of 57 lines (1-57)
 .
 .
 1
 2
 .     total of 56 lines (58-113)
 .
 2
 .
 .
 .
 .
 .
 .
45
 .     total of 101 lines (2528-2628)
 .
45
END
```

57.480	2628	0
61.900	2628	1
67.850	2628	2
61.640	2628	3

END

WinBUGS code for Model 1

```
### T = total number of obs., 8,585; I = total number of students,
2,628; J = total number of schools, 45 ####

model
{

## Level-1 Model ##

for (t in 1: T) {                                        # line 1

 Y[t]~dnorm(mu[t], sigsqinv);                            # line 2

     mu[t] <- pi0[stu[t]] + pi1[stu[t]]*Grade[t];        # line 3
}


## Level-2 Model ##

for (i in 1:I) {                                         # line 4

 pi0[i] ~ dnorm(beta00[sch[i]],tauinv2s[sch[i]]);        # line 5
 pi1[i] ~ dnorm(estudntg[i],tauinv2g[sch[i]]);           # line 6

  estudntg[i] <- betagw[sch[i],1] + betagw[sch[i],2]*(pi0[i]-
                beta00[sch[i]]);                         # line 7

      }

## Level-3 Model ##

for (j in 1:J) {                                         # line 8

beta00[j] ~ dnorm(gamma000,tauinv3s);                    # line 9
betagw[j,1:2]~dmnorm(eschgw[j,1:2],tauinv3gw[1:2,1:2]);# line 10

   eschgw[j,1] <- gamma100 + Bb *(beta00[j]- gamma000);# line 11
   eschgw[j,2] <- Bw0 + Bw1*(beta00[j]- gamma000);      # line 12

        }

# prior specification for fixed effects:

  gamma000 ~ dnorm(0,1.0E-5);                            # line 13
  gamma100 ~ dnorm(0,1.0E-5);                            # line 14
  Bb  ~ dnorm(0,1.0E-5);                                 # line 15
  Bw0 ~ dnorm(0,1.0E-5);                                 # line 16
  Bw1 ~ dnorm(0,1.0E-5);                                 # line 17
```

```
# prior specification for the level-1 precision parameter:

  sigsqinv ~ dpar(1,.0001);                          # line 18

  sigmasq <- 1/sigsqinv;                             # line 19

# prior specification for the level-2 precision parameters:

  for (j in 1:J) {                                   # line 20

  tauinv2s[j] ~ dpar(1,.0001);                       # line 21
  tauinv2g[j] ~ dpar(1,.0001);                       # line 22

  tauvar2s[j] <- 1/tauinv2s[j];                      # line 23
  tauvar2g[j] <- 1/tauinv2g[j];                      # line 24
         }
# prior specification for the level-3 precision parameters:

  tauinv3s ~ dpar(1,.0001);                          # line 25

  tauvar3s <- 1/tauinv3s;                            # line 26

  tauinv3gw[1:2, 1:2] ~ dwish(S[1:2, 1:2],  3);      # line 27
  S[1,1] <-   2.57;                                  # line 28
  S[2,2] <-   0.005;                                 # line 29
  S[1,2] <-   0.000;                                 # line 30
  S[2,1] <-   0.000;                                 # line 31

  tauvar3gw[1:2, 1:2] <- inverse(tauinv3gw[,]);      # line 32
}
```

The first part of the above program (lines 1-3) corresponds to our level-1 model (see Equation 1 above). There are T=8585 time series observations in all, and line 1 essentially sets up a loop in which "t" is an index that runs from 1 to T, thus enabling us to reference each observation. Lines 2 and 3 help us specify Equation 1 in our model. As can be seen in line 2, each observation is assumed normally distributed with expected value "mu[t]" and precision "sigsqinv". In line 3 "mu[t]" is set equal to a value that is a function of the initial status parameter and growth rate parameter for the student ("stu[t]") that the given time series observation is nested in, and the particular grade in which the observation was obtained.

To help grasp how "stu[t]" links the time series data for a given student to that student's growth parameters, note first that "pi0[]" is an $I \times 1$ vector, where I=2628 is the number of students in our sample. The first element of "pi0[]" corresponds to the initial status parameter for student 1, the second element corresponds to the initial status parameter for student 2, and so forth. Similarly, "pi1[]" is an $I \times 1$ vector of growth rate parameters for the students in the sample.

Now let's consider the first time series observation (t=1) in the data set (see the level-1 file). As can be seen math achievement is equal to 70.04, and grade is equal to 0. For this observation, “stu[t]” takes on a value of 1, and so “pi0[stu[t]]” references the initial status parameter for student 1, and “pi1[stu[t]]” references the growth rate for student 1. For the second, third and fourth time series observations we see that “stu[t]” is also equal to a value of 1, and so again “pi0[stu[t]]” and “pi1[stu[t]]” would reference, respectively, the initial status and growth rate parameters for student 1. When we move to the fifth time series observation in the data set, “stu[t]” takes on a value of 2, and so “pi0[stu[t]]” and “pi1[stu[t]]” would reference, respectively, the initial status and growth rate parameters for student 2.

We then specify the level-2 model (lines 4-7). In line 4 we set up a loop in which “i” is an index that runs from 1 to I=2628, thus enabling us to reference each of the students in the sample. Line 5 essentially corresponds to Equation 2a of our model, i.e., the initial status parameter for a given student is assumed normally distributed with mean “beta00[sch[i]]” and precision “tauinv2s[sch[i]]”. (In terms of our naming conventions, note that in the case of the vector “tauinv2s”, the number “2” refers to level 2, and “s” to initial status; hence “tauinv2s” denotes level-2 precisions (variability) connected with initial status parameters.)

Let’s consider the indexing role that “schi[i]” plays here. Note that “beta00[]” is a vector containing the mean initial status parameters for the 45 schools in the sample (i.e., β_{00j} ($j = 1, \ldots, J$)). Thus, for example, let’s focus on the first student in our data set (see the level-2 file above). For this student, “sch[i]” takes on a value of 1, and so “beta00[sch[i]]” would reference the mean initial status parameter for school 1. The same would apply to students 2 - 57. For the 58th student, however, “sch[i]” takes on a value of 2, and so “beta00[sch[i]]” would reference the mean initial status parameter for school 2. The same logic applies in referencing the appropriate precision parameter (i.e., the inverse of the random effects variance in initial status in each school). For the first student in our data set, “tauinv2s[sch[i]]” would reference the initial status precision parameter for students in school 1.

Lines 6 and 7 help us represent the level-2 equation for student growth rates (see Equation 2b above). We first see in line 6 that the growth rate for a given student is assumed normally distributed with an expected rate “estudntg[i]” and precision “tauinv2g[sch[i]]”. (The letter “e” in “estudntg” is used to denote expected values, and the letter “g” in “estudntg” and “tauinv2g” is used to refer, respectively, to expected values and precisions in growth rates.)

Line 7, which we’ll now unpack, sets the expected rate for student i equal to a value that depends on the initial status parameter for student i (“pi0[i]”), and on the initial status/growth rate slope and mean growth rate for the school that student i is nested in. Specifically, “betagw[]” is a J × 2 matrix that consists of 45 rows that contain the school mean growth rates (the β_{10j} ‘s) and within-school initial status / growth rate slopes (the Bw_j ‘s) for the sample of 45 schools. The first row contains the mean growth rate parameter and initial status / growth rate slope for the first school in our sample, the second row contains the mean growth rate parameter and initial status / growth rate slope for the second school, and so on. The use of the index “sch[i]” assures that the each student will be linked to the mean growth rate and initial status / growth rate slope for his or her school. Thus focusing on the first student in our data set (see the level-2 file above), “sch[i]” takes on a value of 1, and so “betagw[sch[i],1]” would reference the mean growth rate for school 1, and “betagw[sch[i],2]” would reference the initial status / growth rate slope for school 1. The same would apply to students 2 - 57. Note, similarly, that “sch[i]” helps us reference the appropriate precision parameter (i.e., the inverse of the amount of random effects variance in growth rates in each school). For the first student in our data set, “tauinv2g[sch[i]]” would reference the growth rate precision parameter for students in school 1.

We then specify the level-3 model (lines 8-12). In line 8 we set up a loop in which “j” is an index that runs from 1 to J=45, thus enabling us to reference each of the schools in the sample. Line 9 essentially corresponds to Equation 3a of our model, i.e., the initial status parameter for school j is assumed normally distributed with mean “gamma000” and precision “tauinv3s”.

In line 10, school mean growth rate and the initial status / rate of change slope for school j (“betagw[j,1:2]”) are assumed multivariate normally distributed with expected values “eschgw[j,1:2]” and precision matrix “tauinv3gw[1:2,1:2]”. (Note that a multivariate normal distribution is specified using the key word “dmnorm” as opposed to “dnorm”.)

Line 10 in combination with lines 11 and 12 helps us represent Equations 3b and 3c. Line 11 sets the expected growth rate for school j (“eschgw[j,1]”) equal to a value that depends on the mean initial status parameter for school j (“beta00[j]”). Similar to Equation 3b, “gamma100” represents the grand mean growth rate, and “Bb” is the latent variable regression coefficient relating school mean initial status (“beta00[j]”) to school mean rates of change. We also see that “beta00[j]” is deviated around the grand mean for initial status (“gamma000”).

We now turn to Line 12 which sets the expected initial status / rate of change slope for school j (“eschgw[j,2]”) equal to a value that depends on the mean initial status parameter for school j (“beta00[j]”). Similar to Equation 3c, “Bw0” represents the average initial status / growth rate slope, and “Bw1” is the latent variable regression coefficient relating school mean initial status (“beta00[j]”) to school initial status / growth rate slopes. In line 11, we see that “beta00[j]” is deviated around the grand mean for initial status (“gamma000”).

In lines 13-17, we specify diffuse priors for the fixed effects in our model, i.e., normal priors with a mean of 0 and extremely low precision.

In line 18, we place a diffuse prior on the level-1 precision parameter “sigsqinv”, i.e., we employ a Pareto prior as described in our paper. We also employ similar priors for the level-2 initial status precision parameters (“tauinv2s[j]”) and growth rate precision parameters (“tauinv2g[j]”) for the 45 schools in our sample (see lines 20-22), where the index “j” runs from 1 to 45. In addition, we specify a Pareto prior for the level-3 initial status precision parameter (“tauinv3s”) (line 25).

In lines 27-31, we specify a diffuse data-dependent Wishart prior for the precision matrix “tauinv3gw[1:2,1:2]”, with scale matrix “S” and 3 degrees of freedom. (See endnote 1 in the text of our article for specific details.)

In each iteration, WinBUGS will generate a value for each of the parameters in the joint posterior distribution based on our model (i.e., the fixed effect parameters, the precision parameters, the initial status and growth rate parameters for the individuals in our sample; and the initial status, growth rate, and initial status / rate of change slopes for the schools in our sample). It is profitable to think of WinBUGS as creating an MCMC data set consisting of a column for each of the parameters in the joint posterior, and the columns are filled with values generated for these parameters over the course of many iterations of the Gibbs sampler.

Suppose that rather than drawing inferences concerning a particular precision parameter, we wish to draw inferences concerning the corresponding variance parameter. This is easily accomplished in WinBUGS. For example, in each iteration of the Gibbs sampler, the statement in line 19 instructs WinBUGS to take the most recently sampled value for "sigsqinv", invert it, and then assign it to the logical node "sigmasq". These values are essentially stored in a new column (vector) of our MCMC data set labeled "sigmasq". Similarly, lines 23, 24, 25 and line 32, instruct WinBUGS to invert values generated for the level-2 and level-3 precisions, respectively, to obtain values for the level-2 and level-3 variances.

<u>Initial Values</u>

As in all MCMC applications, efforts should be made to specify good initial values for the parameters in one's model. There are a number of ways of obtaining good initial values in LVR-HM settings, and we discuss various strategies and possibilities in the context of Model 1.

Level-2 Parameters: Since primary interest in our analyses centers on within-school initial status / rate of change slopes (i.e., the Bw_j's), we first fit the two-level HM specified in Equations 1, 2a and 2b in our JEBS paper to each school's data using HLM6, and used the latent variable regression option in HLM6 to obtain an estimate of Bw_j for each school. These estimates can be used to construct various useful exploratory plots. Furthermore, they provide good initial values for the Bw_j (i.e., "betagw[j, 2]"). In addition, the school-by-school HLM6 analyses provide good initial values for the school mean growth rate parameters ("betagw[j, 1]") and school mean initial status parameters ("beta00[j]").

Note also that estimates of the variance components from the above school-by-school analyses can be used to obtain good initial values for the within-school precision parameters for initial status ("tauinv2s[j]") and growth rates ("tauinv2g[j]").

Level-3 Parameters: Focusing on the level-3 equation for the Bw_j's, a good initial value for the average initial status / rate of change slope ("Bw0") can be obtained by computing the average of the initial values for the Bw_j's. Furthermore, regressing the initial values for the Bw_j's on the initial values for school mean initial status yields a reasonable initial value for the fixed effect capturing how differences in school mean initial status relate to differences in the magnitude of within-school initial status / rate of change slopes ("Bw1"). The estimate of the residual variance from this regression can be used to obtain an initial value for the precision parameter connected with this level-3 equation (i.e., "tauinv3gw[2,2]").

For initial values for the fixed effects and precision parameters in the level-3 equations in which school-mean initial status and school-mean growth rate parameters are treated as outcomes, one strategy is to use HLM6 to fit a standard three-level HM to the data that has the same level-1 model defined in Equation 1 above, but that does not contain any predictors at levels 2 and 3, and use HLM's latent variable option to obtain an estimate of "Bb" (i.e., the latent variable regression coefficient capturing the relationship between school-mean initial status and school-mean rates of change). The resulting estimates could also be used to obtain initial values for the grand mean for initial status ("gamma000"), the grand mean growth rate ("gamma100"), and the level-3 precision parameters connected with school-mean initial status ("tauinv3s") and school-mean growth rates ("tauvar3gw[1,1]").

Level-1 Parameters: Turning to the parameters in the level-1 model, note that when we fit the standard three-level model described above via HLM, the resulting estimate of the level-1 variance can be used to compute an initial value for the level-1 precision parameter ("sigsqinv"). In addition, the resulting level-2 residual file can be used to compute initial values for the initial status and slope parameters for the students in the sample (i.e., "pi0[i]" and "pi1[i]"). Specifying initial values for the individual growth parameters can be quite arduous in this application since there are approximately 2,600 students in the sample. A strategy that simplifies this task somewhat and that has worked well for us in various applications including this one, entails setting the initial values for the initial status and growth rate parameters for each student in a given school equal to the school mean initial status and school mean growth rate initial values for that school.

Note that there are a number of other options and variations on the above strategy. For example, the average of the time 1 math achievement scores for each of the schools in the sample ($\overline{Y}_{1j}$) could be used as initial values for the school mean initial status parameters ("beta00[j]"), and the grand mean of the time 1 math achievement scores could be used as an initial value for "gamma000". Furthermore, we might, for example, regress initial values for the within-school initial status / rate of change slopes and for the school mean growth rates on $\overline{Y}_{1j}$ to obtain initial values for "Bw1" and "Bb", respectively. It should also be noted that the resulting level-3 residual file from the aforementioned three-level HLM6 analysis could be used to compute initial values for the school mean initial status and growth rate parameters for the schools in the sample (i.e., "beta00[j]" and "betagw[j, 1]").

These general ideas and strategies for specifying initial values can be extended to models such as Model 2, which include observed predictors at both levels 2 and 3, in addition to latent variable predictors.