Advice to users on using weights when fitting multilevel (random effects) models

There are no optimal solutions to fitting multilevel models to survey data, but some work better than others. Two-level models where the higher level corresponds to clusters in the sample design are the only models supported by developed theory. Other models can only be based on intuition and we recommend caution.

Suppose you are fitting a two-level linear model with individuals at level 1 and level 2 corresponds to household. Households correspond to the units in the penultimate sampling stage in the general population.

For some users, it could be tempting at this stage to appeal to the notion that one does not need to worry about the survey weights because clustering – an important aspect of the sampling design – has already been accounted for in the model. However, this line of argument is not recommended because other aspects of the design (e.g., stratification, clustering in the primary sampling units (PSU) at the postcode sector level, variation in selection probabilities) have not been included.

Generally, unless you are sure that **every** aspect of the complex sampling design has been included in the model then the survey weights should always be used.

Therefore, we recommend you fit this model using 'pseudolikelihood' estimation (Pfeffermann et al. 1998; Rabe-Hesketh and Skrondal 2006). Pseudolikelihood combines information about the complex sampling design of Understanding Society (UKHLS) and the modelling assumptions implicit in the two-level model to ensure the correct estimates of the model parameters and the standard errors of these parameters are reported.

There is, however, a complication for the pseudolikelihood estimation of two-level models when the analyst is interested in estimates of the variance components (e.g., the variance of the random effects). The overall survey weight must be split into its two components: Level 1 weight $w_{i|h}$ and Level 2 weight w_h where i represents each individual in cluster (household) h. As the overall survey weight is $w_{ih} = w_h \times w_{i|h}$, the user needs to know either,

(i) Level 2 weight w_h and level 1 weight $w_{i|h}$ or
(ii) Level 2 weight w_h and overall weight w_{ih}.

UKHLS provides data users with (ii) as set out in the example (see Box 1).[1]

If you are using Stata, the easiest way to fit your model using pseudolikelihood is to use the **`mixed`** command. This must be done using the maximum likelihood **not** restricted maximum likelihood (REML) option because pseudolikelihood does not work for REML (at least as it is implemented here).

You do not need to set up the survey function using `svyset`. Instead, you set the following options when using `mixed`:

a) Your Level 1 weight in `[pw = <name of level 1 weight var here>]`
b) Your Level 2 weight in `pweight(<name of level 2 weight var here>)`

[1] The two types are closely related: type (i) weights can be created from type (ii) weights simply by dividing the overall weight by the level 2 weight for every individual. If you have type (ii) weight, it is recommended you derive and use type (i) weights. In the case of UKHLS, $w_{i|h}$ should be derived as the released individual weight divided by the released household weight.

c) Make your standard errors robust to PSU clustering by setting `vce(cluster <PSU variable here>)`

As noted above, make sure the `mle` fitting option is specified and not `reml`.

It is recommended that you conduct your analysis using **scaled weights** as set by the `pwscale()` option. Estimating the model without this option runs the risk of producing biased estimates of the variance components (e.g., the random effect variances).

Three scaling options are available each of which promises to reduce bias most effectively in different theoretical scenarios. These options are as follows:

1. `pwscale(size)` – scales the level 1 weights to equal the sample size.
2. `pwscale(effective)` – scale the level 1 weights to equal the *effective* sample size (that is, the size the sample would be if it were drawn using simple random sampling)
3. `pwscale(gk)` – scales the level 1 weights to be equal and the level 2 weight by the mean level 1 weight.

Although we draw a distinction between type (i) and type (ii) weights, this makes no difference to either option 1 or option 2, which will return identical results. However, `pwscale(gk)` requires the type (i) weights to produce the correct results.

We advise estimating the model using all three scaling options to assess the robustness of your variance components estimates to different choices. All (including the unscaled estimates) should perform similarly (but not identically) in terms of the regression coefficients. In terms of the variance components, the theoretical justification for 1-2 relies on the population size in each Level 2 unit being large, but both have been shown to perform well even when this does not hold; conversely, scaling 3 does not make this assumption but is based on an approximation. Korn and Graubard (2003, table 1) show all three sets of weights are shown to have comparable performance for small population clusters.

Finally, weights are not the whole story. If Level 2 of the model does not correspond to the postcode sector PSUs used by UKHLS, cluster-robust variance estimation must be used to account for clustering (see point c) above). However, it should be noted that the effects of stratification are not taken into account so inferences may be slightly conservative.

For MLwiN users, teaching materials about how to use survey weights are available from the Centre for Multilevel Modelling web site (Pillinger 2011). For three-level models and more generally, the weight scaling is more complicated (e.g., Rabe-Hesketh and Skrondal 2006). MLwiN can be run from Stata.

Box 1
Example using Understanding Society data

Assume PSU as level 1, household as level 2 and individual as level 3.

```
global ms “where UKHLS data is stored”

use "$ms/a_indresp", clear
isvar pidp a_hidp a_age_dv a_jbstat a_mastat_dv a_nchild_dv a_scghq1_dv
a_indscus_xw
keep `r(varlist)'
merge m:1 a_hidp using "$ms/a_hhresp", keepus(a_hhdenus_xw) nogen keep(3)
merge 1:1 pidp using "$ms/xwavedat", keepus(psu strata sex_dv ethn_dv
psnenus_xd) nogen keep(3)

mvdecode _all, mv(-21/-1)
recode sex_dv 0=.

global fe_eq1 i.sex_dv c.a_age_dv##c.a_age_dv i.ethn_dv i.a_jbstat
i.a_mastat c.a_nchild_dv

drop if a_hhdenus_xw==0
drop if a_indscus_xw==0

generat wgtlevel1=psnenus_xd
generat wgtlevel2= a_indscus_xw/wgtlevel1

mixed a_scghq1_dv $fe_eq1 [pw=wgtlevel2] || a_hidp:, mle
pweight(wgtlevel1) vce(cluster a_psu) pwscale(gk)
```

References

Korn E, Graubard G. (2003). Estimating variance components by using survey data. J. R Statist. Soc. B **65** 175-190

Pfeffermann D, Skinner C, Holmes D, Goldstein H, Rasbash J. (1998). Weighting for unequal selection probabilities in multilevel models. J. R. Statist. Soc. B **60** 23-40.

Pillinger R. (2011) Weighting in MLwiN. Centre for Multilevel Modelling Learning Materials https://www.bristol.ac.uk/cmm/team/pillinger-learning-mats.html

Rabe-Hesketh S, Skrondal A. (2006). Multilevel modelling of complex survey data. J. R. Statist. Soc. A **169** 805-827