

Weights for multi-level modelling with WERS 2011

John Forth (NIESR)

26th September 2017

The linked SEQ-MQ data from WERS 2011 provides opportunities for multi-level analysis in which the analyst seeks to explore the role of employee and workplace-level characteristics in explaining employee-level outcomes.

This background note accompanies syntax to generate weights for use in multi-level modelling of linked employer-employee data from the 2011 WERS. The syntax itself can be downloaded from the [UK Data Service catalogue page for WERS 2011](#) (scroll to the bottom of the page for the section headed 'Syntax/code').

In the discussion that follows, we employ the terminology and notation used in the 'Survey data' section on pp.323-327 of the [Stata Multilevel Mixed Effects Reference Manual](#).

For a general introduction to multi-level models, see the introductory material provided by the University of Bristol's [Centre for Multilevel Modelling](#).

Weights for multi-level modelling

Multi-level regression estimators (such as Stata's `-mixed-` command) require the user to specify separate weights for each level of the sample design. Specifically, in a two-level employer-employee sample design such as WERS, one is required to specify:

- a Level 2 weight which takes account of the workplace's probability of participating in the SEQ (w_j)
- a Level 1 weight which takes account of the employee's probability of selection, conditional on their workplace having participated in the SEQ (w_{ij}).

The WERS 2011 SEQ dataset contains only a single weight (`seqwtncr_apr13`) which takes account of sample selection at both the workplace level (Level 2) and employee level (Level 1). In other words, the SEQ dataset contains an overall weight w_{ij} which is the product of the two components described above:

$$W_{ij} = w_j * w_{ij}$$

It is therefore necessary to decompose this combined SEQ weight variable (`seqwtncr_apr13`) into its constituent components (w_j and w_{ij}).

Deriving the Level 1 weight (the employee-level component or w_{ij})

If a workplace has been selected for WERS, then the selection process for the SEQ selects employees at random from within the total workforce, where the size of the workforce is given by the WERS MQ variable `ZALLEMPS`. If the workplace has 25 or fewer employees, all employees are selected to take part in the SEQ (a selection probability of 1). If the

workplace has more than 25 employees, then only 25 are selected at random (a selection probability of 25/ZALLEMPS).¹ The selection probability is therefore:

$$\frac{\min(25, \text{zallemps})}{\text{zallemps}}$$

The Level 1 weight ($w_{i|j}$) will be the inverse of this probability, hence:

$$w_{i|j} = \frac{\text{zallemps}}{\min(25, \text{zallemps})}$$

Deriving the Level 2 weight (the workplace-level component or w_j)

All workplaces that participated in the MQ were eligible for the SEQ, so we could use the workplace weight (`estwtnc_apr13`). However, it was apparent that there were some observable non-response biases within the sub-sample of workplaces who eventually participated in the SEQ.² It is thus preferable that we derive a Level 2 weight which corrects for these observed biases. One might imagine that, since that final SEQ weight is the product of the Level 2 weight that we wish to derive and the Level 1 weight that we have derived above, then:

$$w_j = \frac{\text{seqwtnc_apr13}}{w_{i|j}}$$

However, the Level 2 weight must be constant across all employees within the same workplace (it must not vary across Level 1 units who belong to the same Level 2 group). The derivation given above does not meet this requirement, since the final SEQ weight (`seqwtnc_apr13`) includes a small degree of post-stratification by gender. In other words, the weight varies slightly between men and women in the same workplace. For this reason, it is necessary to take the mean of `seqwtnc_apr13` within each workplace and, thus, for each workplace to derive:

$$w_j = \frac{\text{mean}(\text{seqwtnc_apr13})}{w_{i|j}}$$

The post-stratification is thus (incorrectly) incorporated into the Level 2 weight, whereas it should ideally be part of the Level 1 weight. However, the extent of the post-stratification was small and so we treat this as ignorable.

Application

The accompanying syntax gives a simple example of how these Level 1 and Level 2 weights would be called from within Stata's `-mixed-` command.

¹ See p.20-21 of the document 'The Design and Administration of the 2011 Workplace Employment Relations Survey' (supplied with the WERS 2011 data).

² See p.20 of 'The Design and Administration...'

Further information

For further information on any aspect of the 2011 WERS, please consult our website at www.wers2011.info.

If you cannot find the information or advice you are looking for, please use the contact form on the website to submit a query to the WERS team. We endeavour to reply within 48 hours.

John Forth (NIESR)

26th September 2017