

Negotiating the Life Course, Wave 3, Weighting Variables

Working paper only - not for citation

Trevor Breusch *

Negotiating the Life Course Discussion Paper Series

Discussion Paper DP- 020

September 2004

Sampling Weights for Income Units

The method for weighting the income units is documented in detail in the companion paper, “Negotiating the Lifecourse, Wave 1 (1997), Sampling Weights for Income Units and Persons”, dated June 2003. An alternative reference is the consolidated document covering the first two waves, “Negotiating the Lifecourse, Waves 1 and 2: Sampling Weights for Persons and Income Units”, dated July 2003.

Here is a cross-tabulation of the age group of the reference person and the income unit type in the Wave 1 data file for the Wave 3 sample:

Table 1: ‘agegp’ by ‘iutype’, Wave 3

age group of ref person	income unit type (child dep to 18)				Total
	couple wi	couple on	sole pare	single pe	
18-29	47	46	7	126	226
30-44	379	54	59	102	594
45-54	146	93	26	58	323
over 55	9	40	0	0	49
Total	581	233	92	286	1,192

The principle used in constructing the weights is to make the set of income units in the Wave 3 sample representative of the population at the time of the initial sampling in Wave 1. Thus the only difference in the method from the Wave 1 weights is a restriction to the sample of 1192 responses in Wave 3. The weighting adjustment allows for under- or overrepresentation of groups in the original sampling at Wave 1, as well as for different rates of attrition between Waves 1 and 3 for the groups.

Table 2: Weights for Income Units, Wave 3 (variable name ‘iuwt3’)

iutype	q22	agegp			
		18-29	30-44	45-54	over 55
couple with depende		0.7381	0.5785	0.6432	1.2486
couple only		0.8054	0.9848	0.8770	1.0014
sole parent (with d	female	3.4037	0.8191	0.3716	
	male	1.3016	0.8906	1.1849	
single person	female	2.2112	1.5536	1.2462	
	male	2.9852	1.3802	1.0470	

The pattern of weighting reflects the distortions in sampling of the different groups and is similar to the pattern observed in the previous two waves. In some respects the distortion has become more pronounced with the extra attrition to Wave 3. Here the weights range from 0.37 to 3.40, compared to a range from 0.42 to 2.44 in Wave 1 and 0.38 to 2.32 in Wave 2. The biggest change from Wave 2 is in the loss of those females who were young sole parents in the original sample, and of those who were young single persons. This latter effect applies to both sexes, but it is more pronounced with females than males, thus evening up some of the distortions in the sex ratio found in earlier waves. Those who were older couples with dependents have also disappeared since Wave 2 more than other groups, although they are few in number (mostly because of the original sampling scheme). As before, couples who were in the middle age groups with dependents are still over-represented, now by an even greater margin than before. However it is not all bad news: nearly half of the groups are more representative in their proportions in the Wave 3 sample than in Wave 2, although mostly by small margins.

Sampling Weights for Persons

As in Wave 1, the benchmarking for persons is done after the sample has been weighted to correct for the effect of the sample design in which the probability of selection of a person varied inversely with the number of eligible persons in the household. Again the variable ‘hhwt’ is used, with the modification that the largest group in a household is ‘four or more’ eligible persons. The benchmarking method is the same as in Wave 1, but here the sample is restricted to the respondents who are present in Wave 3. The numbers of persons in the weighting categories and the resulting person weights are as follows:

Table 3: ‘agegpr’ by ‘numhh’, Wave 3

agegpr	q22	number eligible in household				
		1	2	3	4 or more	
18-29	female	34	69	13	8	124
	male	19	60	20	9	108
30-44	female	111	229	15	2	357
	male	85	190	9	4	288
45-54	female	78	66	28	10	182
	male	45	61	18	9	133
Total		372	675	103	42	1,192

Table 4: Weights for Persons, Wave 3 (variable ‘pswt3’)

agegpr	q22	number eligible in household			
		1	2	3	4 or more
18-29	female	0.8146	1.6292	2.4437	3.2583
	male	0.8474	1.6948	2.5422	3.3896
30-44	female	0.4190	0.8380	1.2569	1.6759
	male	0.5008	1.0017	1.5025	2.0034
45-54	female	0.4154	0.8308	1.2462	1.6616
	male	0.5491	1.0983	1.6474	2.1966

The range of weights is now 0.42 to 3.39, where it was 0.42 to 2.92 in Wave 2, and 0.44 to 2.70 in Wave 1. Still it is the males in all groups, and the young of both sexes, who are underrepresented, particularly those who were living in larger households. Attrition from Wave 2 is highest in the youngest age group, thus exacerbating the distortions found in earlier waves. There has, however, been some evening up of the sex distributions in the sample. Indeed, half the groups in Table 4 have weights closer to unity than they had in Wave 2.


```

mat R1=(1,1,1,1\1,1,1,1\1,1,1,1\1,1,1,1);
sca sumN1=0;
forval i=1/4 {; forval j=1/4 {;
  sca sumN1=sumN1+N1[`i',`j'];
}; };
forval i=1/3 {; forval j=1/4 {;
  mat R1[`i',`j']=(C1[`i',`j']/sumC1)/(N1[`i',`j']/sumN1);
  replace iuwt3=R1[`i',`j'] if agegp==`i'&iutype==`j';
}; };
* Now to handle the last row which has the zeros;
mat R1[4,1]=(C1[4,1]/sumC1)/(N1[4,1]/sumN1);
mat R1[4,2]=(C1[4,2]/sumC1)/(N1[4,2]/sumN1);
forval i=1/4 {;
  replace iuwt3=R1[4,`i'] if agegp==4&iutype==`i';
};
* Modify these weights for the gender ratio in each agegroup
* for iutype=3 and 4. These iutypes have only 3 age groups;
tab agegp q22 if iutype==3, matcell(N3) row;
tab agegp q22 if iutype==4, matcell(N4) row;
mat list N3;
mat list N4;
* For iutype=3, weight each cell separately (although cells
* have very small counts);
mat R3=(1,1\1,1\1,1);
forval i=1/3 {; forval j=1/2 {;
  mat R3[`i',`j']=(C3[`i',`j']/(C3[`i',1]+C3[`i',2]))/
  (N3[`i',`j']/(N3[`i',1]+N3[`i',2]));
  replace iuwt3=iuwt3*R3[`i',`j'] if agegp==`i'&q22==`j'&iutype==3;
}; };
* For iutype=4, weight each cell sepatately (although cells
* have very small counts);
mat R4=(1,1\1,1\1,1);
forval i=1/3 {; forval j=1/2 {;
  mat R4[`i',`j']=(C4[`i',`j']/(C4[`i',1]+C4[`i',2]))/
  (N4[`i',`j']/(N4[`i',1]+N4[`i',2]));
  replace iuwt3=iuwt3*R4[`i',`j'] if agegp==`i'&q22==`j'&iutype==4;
}; };
* Check there is no rounding error. IU weights should average=1.0000;
summ iuwt3;
label var iuwt3 "weight for income units W3";
*****;
* Weights for Persons, weighting first by hhwtm2 (modified form
* maximum of 4 persons/household). Allows attrition to Wave 3;
*****;
* First rebalance hhwtm to average to 1.0 on the reduced sample;
egen sumhhwtm=sum(hhwtm);
gen hhwtm2=(hhwtm/sumhhwtm)*_N;

```

```

tab agegpr q22 [aw=hhwtm2] if agegpr>=1&agegpr<=3, matcell(N5) row;
mat list N5;

* Population values from Census 1996
* Rows `i' are agegpr=1/3, cols `j' are Q22=1/2;

mat C96=( 1591993, 1582414\
          2033996, 2083361\
          1128244, 1109180);

sca sumC96=0; sca sumN5=0;

forval i=1/3 {; forval j=1/2 {;
    sca sumC96=sumC96+C96[`i',`j'];
    sca sumN5=sumN5+N5[`i',`j'];
}; };

mat R5=(1,1\1,1\1,1);
gen pswt3=.;

forval i=1/3 {; forval j=1/2 {;
    mat R5[`i',`j']=(C96[`i',`j']/sumC96)/
        (N5[`i',`j']/sumN5);
    replace pswt3=hhwtm2*R5[`i',`j'] if agegpr==`i'&q22==`j';
}; };

* Check there is no rounding error. Person weights should average=1.0000;

summ pswt3;
label var pswt3 "weight for persons W3";

*****;
** Finish up;
*****;

keep id iuwt3 pswt3 both;
sort id;
save temp3.dta, replace;

use temp1.dta, clear;
merge id using temp3.dta;
tab _merge;
drop _merge;
sort id;

cor hhwt hhwtm iuwt pswt iuwt3 pswt3;

by iutype, s: tab iuwt3 agegp if q22==1;
by iutype, s: tab iuwt3 agegp if q22==2;
by q22, s: tab pswt3 agegpr;

sort id;
compress;
desc;
save weights3.dta, replace;
log close;

```