

NLC Data Update 2 – Income Update

Changes to the NLC 1997 Working Dataset

This is the second update in the Data Update Series for the Negotiating the Life Course data. These updates are to ensure that NLC data users are aware of any changes made to the data. If you have any queries concerning the following data update or suggestions for changes to the NLC data (new variables you have created or variables that you would like created) please email nlc@anu.edu.au.

This update focuses on work that has been carried out on the income data by Rob Ackland, Trevor Breusch and Deborah Mitchell. Some of the previous income variables have been removed from the working dataset and have been replaced with updated variables.

CHANGES TO EXISTING INCOME VARIABLES IN NLC WAVE 1 (1997)

The following variables have been removed from the working dataset:

NET2A	Final net yearly salary
NETINC23	Net 2A + 23% tax other/businc
NETINC15	Net 2A + 15% tax other/businc

The following variables have been added to the working dataset:

RINC	Income of respondent
PINC	Income of partner
FINC	Income of family (respondent and partner)

ACCESSING THE REVISED DATA

The updated data is available from nlc@anu.edu.au.

Income for Waves 1 and 2 of the Negotiating the Life Course Survey: Further Changes

Trevor Breusch and Deborah Mitchell

September 2002

The starting point for our work is the earlier work by Robert Ackland, documented by him as [“Constructing an Income-Based Measure of Economic Welfare for Waves 1 and 2 of the Negotiating the Life Course Survey”](#), another paper for the NLC Workshop 17-18 May, 2002. The changes made by us are as follows:

1. In Wave 1, where ‘wage’ is constructed by Ackland from an inverse application of tax tables to a response in net income, this variable is now rounded to the nearest whole dollar.
2. The attempt by Ackland to impute the rental value of owner-occupied housing is removed. Our measures of income exclude imputed rent.
3. Child maintenance payments received are now attributed to the individual recipient (respondent or partner). Ackland included all child maintenance receipts in respondent income.
4. The aggregate measures of respondent, partner, and family incomes are now simple sums of their components. The implication is that the aggregate value will be recorded as a missing value (Stata code ‘.’) if any component is missing. Ackland used the Stata command ‘rsum()’ to create his aggregates, which treats missing components as zeros. He also provided an indicator dummy variable to show when missing income components prevent a meaningful aggregate from being formed. This indicator is no longer needed because the missing value code provides the same information.
5. We make many changes to individual cases, where some or all of the income data for that case was reported by Ackland to be missing. We use our judgement to do this work. Many of the changes are simply setting to zero those income components that are missing but which in the scheme of things are probably trivial (like ‘other income’ or ‘child maintenance’) or which are probably zero anyway (when ‘dk’ is the response to Family Allowance amount for a family that is well above the income limits for entitlement). We have also imputed benefits from Social Security tables where the response is ‘dk’ and removed several instances of apparent double counting.

Variables

The variables in our output data file are

id		
wage	wages	
businc	self-employment/business income	
govben	government benefits	
otherinc	other income	
rchmain	child maintenance, r	
pchmain	child maintenance, p	
rinc	income of r	wage+ businc+ govben+ otherinc+ rchmain
pinc	income of p	(incl pchmain)
finc	income of family (r+p)	rinc+pinc

Documentation

The individual changes are documented in, and carried out by, a Stata 7 'do' file for each Wave. In both cases the input to our process is Ackland's output file that was delivered to the NLC Workshop 17-18 May, 2002 (files called inc97.dta and inc00.dta). In the case of Wave 1 we had to recreate his output file (with a different name) so that it included one extra variable, called 'inc_ref', which is created by his processes but dropped from the output file he presented to the Workshop.

There is a difference in our documentation of the changes between the two waves. In both waves the changes themselves are described clearly by the commands in the 'do' files that make the changes. In both waves there is written documentation of the reasons for the changes in a handwritten log kept by Deborah Mitchell. In Wave 2 but not in Wave 1 the reasons for decisions are also included in the 'do' file. It is entered into a string variable which could be included in the output variables if a minor modification was made to the 'do' file.

Wave 1 (1997) outcomes

Of the 2231 responses in the original file, Ackland noted that 27 of them have no income data at all. In a further 33 cases the respondent initially refused to answer the question about wages and salary, and then was not asked further questions about their own sources of income, but later they gave an informative answer to the supplementary question about their total income in a range (q249). Ackland estimated respondent income for these people by a representative value in the range. In 8 of these 33 cases we attribute the income to a source. But in the other 25 cases no information is available to enable the respondent income to be split into wages and salary, business income, government benefits and child maintenance. In these latter cases there is a value for rinc but the components are missing.

In addition to the 27 cases where nothing is available on income and the 33 cases where respondent income is available in total but not in its components (at least not until we make some attributions), there are 227 cases indicated by Ackland for which a component of family income is missing. After our efforts there remain 44 cases with a missing value for respondent income. Some of these cases have values for some components of income (often zeros) which we have left, but we judge that the missing components may well be significant and hence respondent income is unknown and indicated as a missing value. There are 49 cases in which partner income is missing, 38 cases where both respondent and partner income are missing, and hence there are 55 cases where family income is missing because either the respondent or partner value is missing.

Wave 2 (2000) outcomes

In Wave 2 the income questions were more tightly worded and respondents with business incomes were asked to give their income from their income tax returns. This should have the advantage of better distinguishing wages and business income for the self-employed, but the coding of these responses was confusing and needed a certain amount of disentangling. Much of this work of disentangling was done by Ackland (op cit).

Of the 1768 responses in Wave 2, there are 320 case in the Ackland's output file in which some component of income is missing. This is an increase on the 254 cases of partially or totally missing income data in Wave 1. Given the attrition from Wave 1 to Wave 2 it represents a larger potential loss of data (18.1% compared to 11.4% earlier).

The refusal rate for self-employment income in Wave 2 was more than twice that of Wave 1. If 'don't know' is taken as a polite form of refusal, there are 71 refusals on the amount of business income out of 435 respondents (16.3%) who indicated they have some business income in Wave 2, but only 24 refusals out of 339 respondents (7.1%) in Wave 1.

Notice also that many more respondents reported having some business income in Wave 2, probably reflecting the different ordering of the questions between the waves. In Wave 1 the business income questions came after wages and salary questions, and the preamble to the questions on business income was “Excluding and wages or salary income you’ve already told me about ...”. In Wave 2 the business income question came first. The change in the proportions reporting self-employment in their main job is much smaller: 266 of 1438 employed respondents (18.5%) in Wave 2 against 301 of 1742 employed respondents (17.3%) in Wave 1.

Unlike Wave 1, if the respondent refused to answer the question about wages and salary the questions on other components of income were not skipped. However there was no follow-up question in Wave 2 corresponding to q249 in Wave 1 about total respondent income in a range for those who refused on wages and salary.

After our efforts there remain 87 cases with a missing value for respondent income. Some of these cases have values for some components of income (often zeros) which we have left, but we judge that the missing components may well be significant and hence respondent income is unknown and indicated as a missing value. There are 71 cases in which partner income is missing, 58 cases where both respondent and partner income are missing, and hence there are 100 cases where family income is missing because either the respondent or partner value is missing. In absolute terms this is nearly twice the loss rate of Wave 1, and as a percentage of the number of respondents it is more than twice as high.

Wave 1 (1997) variables:

```

id ----- (unlabeled)
      type:  numeric (int)
      range:  [1,2574]
unique values: 2231
      units:  1
      coded missing: 0 / 2231
      mean:   1306.15
      std. dev: 738.201
      percentiles:
                10%    25%    50%    75%    90%
                  280    668    1311   1956   2319

wage ----- wages
      type:  numeric (float)
      range:  [0,600000]
unique values: 420
      units:  1
      coded missing: 57 / 2231
      mean:   25530.6
      std. dev: 29151.4
      percentiles:
                10%    25%    50%    75%    90%
                  0    3000   23000   39000   52000

businc ----- self-employment/business income
      type:  numeric (float)
      range:  [0,250000]
unique values: 87
      units:  1
      coded missing: 62 / 2231
      mean:   3355.22
      std. dev: 15861.2
      percentiles:
                10%    25%    50%    75%    90%
                  0      0      0      0      5000

```

```

govben ----- government benefits
      type: numeric (int)
      range: [0,22308]          units: 1
unique values: 261             coded missing: 62 / 2231
      mean: 2089.57
      std. dev: 3891.45
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0        1820    8320

otherinc ----- other income
      type: numeric (float)
      range: [0,150000]        units: 1
unique values: 104           coded missing: 54 / 2231
      mean: 1277.59
      std. dev: 6217.28
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0         0        2500

rchmain ----- child maintenance, r
      type: numeric (int)
      range: [0,21580]         units: 1
unique values: 31           coded missing: 62 / 2231
      mean: 156.719
      std. dev: 1079.83
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0         0         0

pchmain ----- child maintenance, p
      type: numeric (int)
      range: [0,7800]          units: 10
unique values: 10           coded missing: 63 / 2231
      mean: 15.9502
      std. dev: 262.973
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0         0         0

rinc ----- income of r
      type: numeric (float)
      range: [0,650000]        units: 1
unique values: 1199         coded missing: 44 / 2231
      mean: 32576.2
      std. dev: 32339.8
      percentiles:      10%      25%      50%      75%      90%
                       5000    12950    27040    42500    61770

```

pinc ----- income of p
 type: numeric (float)
 range: [0,112130] units: 1
 unique values: 31 coded missing: 49 / 2231
 mean: 19236.6
 std. dev: 25103.8
 percentiles: 10% 25% 50% 75% 90%
 0 0 7310 33667 46311

finc ----- income of family (r+p)
 type: numeric (float)
 range: [0,683667] units: 1
 unique values: 1646 coded missing: 55 / 2231
 mean: 51976.4
 std. dev: 41339.7
 percentiles: 10% 25% 50% 75% 90%
 12454 25471.5 44517.5 67618.5 97311

Wave 2 (2000) variables:

id -----
 type: numeric (int)
 range: [1,2574] units: 1
 unique values: 1768 coded missing: 0 / 1768
 mean: 1316.13
 std. dev: 735.239
 percentiles: 10% 25% 50% 75% 90%
 279 683.5 1328 1958.5 2313

wage ----- wages
 type: numeric (float)
 range: [0,1248000] units: 1
 unique values: 279 coded missing: 79 / 1768
 mean: 32676.1
 std. dev: 49902.4
 percentiles: 10% 25% 50% 75% 90%
 0 8000 27000 46000 65000

businc ----- self-employment/business income
 type: numeric (float)
 range: [0,2000000] units: 1
 unique values: 124 coded missing: 56 / 1768
 mean: 10378.8
 std. dev: 55858.2
 percentiles: 10% 25% 50% 75% 90%
 0 0 0 0 35000

```

govben ----- government benefits
      type: numeric (int)
      range: [0,25662]          units: 1
unique values: 208             coded missing: 17 / 1768
      mean: 1847.16
      std. dev: 3918.44
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0        1560     8034

otherinc ----- other income
      type: numeric (float)
      range: [0,200000]        units: 1
unique values: 116           coded missing: 38 / 1768
      mean: 1518.91
      std. dev: 7081.55
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0        100     4000

rchmain ----- child maintenance, r
      type: numeric (int)
      range: [0,21528]        units: 1
unique values: 34           coded missing: 11 / 1768
      mean: 178.641
      std. dev: 1292.61
      percentiles:      10%      25%      50%      75%      90%
                       0         0         0         0         0

pchmain ----- child maintenance, p
      type: numeric (int)
      range: [0,2340]         units: 10
unique values: 6            coded missing: 11 / 1768
      tabulation:  Freq.  Value
                   1751   0
                   1     520
                   1    1040
                   1    1560
                   1    1820
                   2    2340

rinc ----- income of r
      type: numeric (float)
      range: [0,2624000]      units: 1
unique values: 933          coded missing: 87 / 1768
      mean: 47060.7
      std. dev: 84032.7
      percentiles:      10%      25%      50%      75%      90%
                       9450     20000     35000     54600     83000

```

```

pinc ----- income of p
      type: numeric (float)
      range: [0,120239]          units: 1
unique values: 28              coded missing: 71 / 1768
      mean: 24407.5
      std. dev: 30303.7
percentiles:    10%    25%    50%    75%    90%
                0      0     13324  40126  64668

```

```

finc ----- income of family (r+p)
      type: numeric (float)
      range: [0,2642654]       units: 1
unique values: 1315           coded missing: 100 / 1768
      mean: 72001.1
      std. dev: 89168.1
percentiles:    10%    25%    50%    75%    90%
                21000  35112  57762  85860  126228

```

Trevor Breusch & Deborah Mitchell
September 2002