A note comparing estimates of headcount poverty rates and Gini coefficients from grouped data with those from individual data

Andy Newell, February, 2013

Between 1905 and 1909 the Labour Department of the Board of Trade conducted studies of living standards in the industrial centres of five advanced manufacturing countries: Belgium, France, Germany, the United Kingdom and the United States of America. The Board's purpose was to inform the Tariff Reform debate with information on comparative living standards. For four of those countries, all that remains of the original data are tables of average income, family size and food expenditures by household income group. As a prelude to analysing these data, in this note I investigate whether there are significant biases in estimating parameters of income distributions from grouped data, compared to estimating from the underlying individual household data. In principle, employing grouped data should not create biases. For the UK data, we can make direct comparisons between the results using the tabled data with results from a subset of original returns that was located and digitised a few years ago - see Gazeley and Newell, (2011), for instance. The Board of Trade's original work on the UK is reported in 1905 [Cd 2337] and 1908 [Cd. 3864]. We utilise Chen, Datt and Ravallion's POVCAL[i] programme for computing income distribution parameters from grouped data. Ravallion and Huppi (1991) illustrate the methods in the programme.

In Table 1 we compare three methods of estimating summary income distribution statistics. The poverty line is Gazeley and Newell's (2011) adaption of Bowley's (Bowley and Burnett-Hurst, 1915) poverty line. It is a poverty line designed to be applied to income net of rent. In order to apply it to the grouped data of Cd 2337, it is necessary to (a) add an estimate of rent to the poverty line and (b) apply an average household poverty line to all households, rather than compare income net of rent to a household specific poverty line. This is one possible source of bias if the number of truly poor households excluded from this poverty count is not offset by a similar number of non-poor households included in the count. The results of this process are given in row 1 of Table 1. We estimate the Gini coefficient to be 18.8% and find 11.3% of households to be in poverty. This very low Gini reflects the relative (to the household population) homogeneity of the Board of Trade survey sample, which contains households of manual industrial workers and oversamples those of skilled workers in particular. See Gazeley and Newell (op. cit.) for a discussion.

In the second and third rows, for comparison, we calculate the same statistics in different ways, using Gazeley and Newell's (op. cit.) subsample of the original 1904 survey data set. This rediscovered subsample contains about 50% of the original households, with a strong bias towards Scotland. This is a second source of bias, which we eliminate by focussing only on Scotland in the lower part of the table. In row 2 we present the more usual estimates based on household level data. Household-specific Bowley poverty lines are used to compare

against household income net of rent. We find 12.1 % of this sample to have been in poverty on this measure. The results are reassuringly similar to those in the first row, though with a little more poverty and inequality.

In row 3 we come to answering the cleaner question which is: if we take the recovered subsample and estimate the poverty rate and the Gini, but this time using data grouped into the same groupings as the Board of Trade, what do we find? Here again a general average household poverty line, rather than a household-specific line, is applied. What we find is similar to row 2 results, though the inequality and poverty measures are a little higher.

In summary, what emerges from these three sets of findings is (a) that in this case grouping the data raises the Gini and poverty measures a little and (b) that the grouped data estimates of inequality and poverty are 10-15%, or between one and three percentage points higher in the recovered sub-sample than in the full, original data set.

We can mostly eliminate the difference between the full sample and the recovered sample if we narrow our focus to Scotland. We cannot completely eliminate that difference. Here is why: we use 476 Scottish observations, which is more than the Board of Trade used, but we don't know which ones the Board of Trade eliminated. In the lower part of the table we show that the headcount and the Gini are slightly higher for us than for the Board of Trade, if we use grouped data. We also show, consistently with the UK results above, that in these data the Gini and poverty measures are a little higher if grouped data, rather than individual household data, are employed.

Table 1 Estimates of sample household income distribution, inequality and poverty measures

	Row	*Mean income net of rent*	*Gini %*	*Household poverty headcount %*	*Sample Size*
Grouped UK data from Cd 2337	1	£1.65	18.8	11.3	1944
UK BoTR direct	2	£1.63	19.7	12.1	914
UK BoTR grouped	3	£1.63	21.9	12.8	914
Grouped Scottish data from Cd 2337	4	£1.65	18.1	8.6	455
Scotland BoTR direct	5	£1.66	18.5	7.6	476
Scotland BoTR grouped	6	£1.66	19.1	9.0	476

Note: see the text of differences in method of estimation.

References

1905 [Cd. 2337] British and foreign trade and industry (second series). Second series of memoranda, statistical tables, and charts prepared in the Board of Trade with reference to various matters bearing on British and foreign trade and industrial conditions

1908 [Cd. 3864] Cost of living of the working classes. Report of an enquiry by the Board of Trade into working class rents, housing and retail prices, together with the standard rates of wages prevailing in certain occupations in the principal industrial towns of the United Kingdom. With an introductory memorandum

Bowley, A.L., and Burnett-Hurst, A.R. (1915), *Livelihood and Poverty*, G. Bell and Sons, Ltd, London.

Gazeley, I. and A. Newell, (2011) 'Poverty in Britain in 1904: An Early Social Survey Rediscovered,' *Economic History Review*. 64, 1, 52-71.

Ravallion, M. and Huppi, M. (1991) "Measuring Changes in Poverty: A Methodological Case Study of Indonesia During an Adjustment Period", *World Bank Economic Review*, January.

[i] POVCAL is available at http://iresearch.worldbank.org/PovcalNet/index.htm?0,5