LOCAL INCOMES AND POVERTY IN SCOTLAND

ANNEX E

Precision Of Estimates

March 2013

Glen Bramley and David Watkins
(Institute for Housing, Urban and Real Estate Research,
School of the Built Environment,
Heriot-Watt University,
Edinburgh EH14 4AS
0131 451 4605
g.bramley@hw.ac.uk)

Users of the estimates of local and small area incomes and poverty generated in this study will have an understandable desire to know how accurate or precise these estimates are. In conventional statistical analyses, based on a single unified dataset, such estimates can usually be provided in the form of standard errors and confidence intervals on the estimates generated by the model. Because the methodology used in this study is a hybrid mixed method entailing three stages (micro models, synthetic small areas forecasts, and controlling) these standard types of measure are not available.

To meet this challenge, we have developed an indirect procedure to demonstrate the effective degree of precision (or error) in the small area estimates. This procedure entails dividing our sample data into 'cells' which have the characteristics of (a) having roughly the same number of household observations within them as a Scottish Datazone, while (b) being relatively homogeneous in terms of their socio-demographic composition and geographical location. Two sets of cells are defined, one appropriate for use within Scotland with our main Scottish data source, the Scottish Household Survey (SHS), and the other appropriate for use across Scotland and England (excluding London) using our main UK data source for small area estimates, Understanding Society (US). These cells are defined on the basis of combinations of small area type (based on the ONS 'Group' classification) and geographical region (US) or local authority type (SHS). Where cells would be too small, some region-group combinations are combined. Where cells would be too large, a further division is made based on LA names.

We can use these cells to compare predicted income levels or poverty rates with actual levels/rates for sample households, averaged across the individual households within the sample in each cell. Two types of comparison are possible. Firstly, we can compare the predictions from the micro-models with the actual levels for the sample. This gives a measure of the intrinsic precision (or error) of the micro model at this level. Secondly, we can compare the controlled predictions from the synthetic

second stage models with the actuals, for the same set of cells. We make these comparisons for both the SHS-based models and the US-based models.

Although these comparisons are made for both the SHS-based models and the US-based models, some income/poverty measures are only available from US, and some of the SHS predictions are not replicated using US, so not all measures are available from both sources. The main measure of (im)precision is the standard deviation of the 'error' (proportional difference between actual and predicted). A second measure used is the correlation coefficient between actual and predicted values at the cell level. Table E.1 shows the comparisons for the micro model predictions, illustrating the intrinsic imprecision of the first stage model.

Table E.1: Errors and Correlations using Micro Predicted Values versus Actual Survey Values at 'Cell' Level

Intrinsic Micro pred vs act		*SHS Std Dev of Errors*	*US Std Dev of Errors*	*SHS Correlns*	*US Correlns*
Gross Hshld Income	errginca	0.078	0.057		0.958
Net Hshld Income	errninca	0.050	0.048	0.985	0.958
1st Benefit Unit Income	errbinca	0.078			
Net Equiv BHC	errneiba	0.039	0.062	0.978	0.959
Net Equiv AHC	errneiaa	0.043	0.074		0.955
Poverty BHC	errpovba	0.100	0.118	0.937	0.919
Poverty AHC	errpovaa	0.098	0.129		0.912
Material Depriv	errmd4a		0.133		0.948
Pov + Mat Dep	errpovmda		0.311		0.942

The table shows that the standard deviation (i.e. the average variation around zero) of the proportional prediction errors of the SHS micro model is 0.078 (i.e. 7.8%) for gross household income level and for FBU income, while a lower proportional errors apply to net household income (0.050, or 5%) and to net equivalent income BHC (0.039) and AHC (0.043). The US-based models achieve somewhat lower errors in the case of gross (0.057) and net income (0.048) but rather higher errors for equivalent income (0.062 and 0.074). The predicted poverty rates are less precise in proportional terms (0.100 and 0.098 in SHS, 0.118 and 0.129 in US). However, it should be noted that these are not the same as the errors in terms of percentage points. For example, an average proportional error of 0.10 on a poverty rate whose mean value is 20% would be 2 percentage points. The error in the US prediction of material deprivation is similar to that for low income AHC, at 0.133. However, for the combined poverty and material deprivation measure, which varies more sharply in proportional terms, the proportional error is higher at 0.311.

The correlations shown paint a generally positive picture of a fairly high degree of correspondence between the predicted and actual values at cell level. In SHS these

are around 0.98 for average income levels and around 0.94 for poverty rates. In US these are around 0.96 in the former case and 0.92-0.95 for the poverty measures.

Table E.2 shows the proportional errors when comparing synthetic stage 2 model predictions (controlled) with the actual values at cell level. We show SHS and US separately and the final stage combined and controlled values in the third column. While the synthetic models may introduce additional sources of error, relating to the aggregation of process and detailed differences in some of the variables used, this is offset by the controlling process, so these are generally similar or a bit more accurate than the stage 1 model comparisons.

Table E.2: Errors and Correlations using Controlled Synthetic Predicted Values versus Actual Survey Values at 'Cell' Level

Final Synthetic Controlled vs Act		*SHS Std Dev of Errors*	*US Std Dev of Errors*	*Combined Std Dev of Errors*
Gross Hshld Income	errginca	0.039	0.062	
Net Hshld Income	errninca	0.040	0.051	
1st Benefit Unit Income	errbinca	0.041		
Net Equiv BHC	errneiba	0.032	0.045	0.045
Net Equiv AHC	errneiaa	0.051	0.045	0.045
Poverty BHC	errpovba	0.117	0.120	0.119
Poverty AHC	errpovaa	0.089	0.128	0.127
Material Depriv	errmd4a		0.157	0.188
Pov + Mat Dep	errpovmda		0.250	0.379

These comparisons show that the SHS-based predictions come closer to the actuals in the cases of gross and net income, FBU income, and AHC poverty. The average net equivalised income AHC is slightly less close, as is the BHC poverty incidence. Using US, the predictions are slightly less close except in the case of equivalised income AHC – this may reflect weaknesses in the SHS measurement of housing costs. The combined predictions used for some indicators are similar to the US-based ones, except that the errors are rather greater for material deprivation and combined poverty.

Assuming a normal distribution of errors, one can say that the 95% confidence interval on the cell-level predictions will be plus or minus twice the standard error (i.e. the standard deviation of the errors). Assuming again that our sample cells are a good proxy representation of datazones, then we would conclude that the confidence intervals for these estimates would be as follows:

- For median total household income, net income or BHC income, approximately 8-10%
- For net equivalent incomes, approximately 9%

- For low income poverty incidence BHC or AHC, approximately 18-25% of the mean value of poverty incidence, which is equivalent to 3.5-5 percentage points in the estimated poverty incidence rate.

- For material deprivation incidence, approximately 31-38% of the mean value for this indicator, which is equivalent to 5.3-6.5 percentage points in the estimated material deprivation rate.

- For combined poverty and material deprivation, approximately 50-75% of the mean incidence value, equivalent to 3.5-5.3% percentage points in the estimated incidence rate