

Income Imputation in the SHS

Final Report – February 2010

Part 1 – Background and Research

Phil Clarke, ONS

1. Background

1.1. ONS has undertaken a Quality Improvement Funded project for Scottish Government aimed at improving local authority estimation of household poverty.

1.2. **Poverty measurement** Poverty estimation is measured as proportion of households whose equivalised income is below a threshold value (poverty line). Income is equivalised to allow households of different sizes to be objectively compared. There are two measures – before and after taking housing costs into account. In this study, we look at the before housing costs measure. Additionally there are separate concepts of absolute and relative poverty. Absolute poverty uses a value of the poverty line which is fixed over different years. However, it is relative poverty which is of interest here. In this case the poverty line is defined each year as 60% of median national household income. The national here refers to that for Scotland as a whole.

1.3. The Scottish Household Survey (SHS) is a large survey which includes questions on income and is of a suitable size – over a two year period - for local authority estimation. However it suffers from a shortcoming for poverty estimation in that detailed income questions are only asked of the main household respondent – the Household Reference Person (HRP) and his/her spouse if present. This means that a full household income (either total or equivalised) cannot be ascertained if there are other adults in the household.

1.4. The Family Resources Survey is a detailed household survey conducted by DWP and is used for poverty estimation throughout the UK. This survey does consider the income of each member of surveyed households, however it is not sufficiently detailed for local authority estimation.

1.5. Imputation of income for the missing adults in SHS households has been proposed as a solution. With this additional data, full and equivalised household incomes can be determined. Hence it becomes possible to determine extent of poverty estimation by local authority. As it does cover the missing adults, a source for the imputation could be the FRS.

1.6. Ipsos MORI who are contractors to Scottish Government in conducting the SHS have carried out a study of the requirement. They have issued a report proposing a strategy of hierarchical hot deck imputation from FRS to achieve the poverty estimates. This is summarised below. Their data analysis is based on 2005/06 FRS and SHS.

Summary of Ipsos MORI's proposed strategy

1.7. Principle of imputation from FRS (Sections 1 & 2 MORI report)

MORI examined the comparability of the income collected by the SHS and FRS in terms of both definitional content and respondent coverage. The content difference identified was that certain benefits were included in income variables for the FRS but not SHS -:

- Lone parent benefit run-on;
- War widows pension;
- Maternity grant, funeral grant or community grant from Social Fund;
- Guardians allowance;
- Work search premium;
- Bereavement payment lump sum;
- Winter fuel payment lump sum;

Back to work bonus;
Child maintenance bonus;
Government training allowances.

All of these are of minor consequence apart from winter fuel payments for pensioner households.

The respondent coverage difference was, as noted above, that FRS collects from all adults in a household plus also income by children (i.e. all benefit units). SHS collects data only from the household reference person (HRP) and spouse if present (benefit unit 1). While income by children is negligible, income from other adults would be needed for total household income.

Table 2.2 (MORI report) shows for each household composition type, the percentage of FRS reported income by HRP/spouse/other and by SHS inclusion definition. Only the first two columns represent income that will be reported in the SHS. It is clear that the only substantial elements of difference lie in the column headed “Others – SHS def” occurring only in rows for three or more adults with children and “Two adults, no children one over pension age” – this latter would predominantly represent adults of different generations who would be different benefit units and only the HRP would be interviewed in the SHS.

The conclusion drawn was that a poverty estimation from SHS using imputation from FRS source for those household members omitted would be feasible. To make income directly comparable to SHS, components collected only in the FRS could be deducted.

Because of the deduction, poverty estimates could not be directly compared with those published from FRS for Scotland, However, a further reason for non-comparability here is that the FRS poverty estimates are based on a Great Britain rather than a Scotland poverty line, i.e. 60% median equivalised income over England, Scotland and Wales.

1.8. Determining imputation strategy (Section 3 MORI report)

1.8.1 In section 3, MORI describe various possible imputation strategies, both deterministic and stochastic. Deterministic strategies are felt to suffer from a major shortcoming, that they would reduce the variability of the data and are thus not suitable. Three stochastic methods are described.

The first stochastic method described involves modelling the FRS data to examine the differences between income for HRPs and other adults. These ratio differences could then be applied to SHS records for HRPs and donated in a random manner to other adults.

The second stochastic method described involved using the “random adult” from the SHS as a source of imputations. The random adult is an additional response record in the SHS who could be any adult in a household – HRP, spouse or other adult – and which would hold income information.

The third stochastic method would be to model income (after deducting components not in the SHS definition) from the FRS for other adults against a set of determining variables. Groups of donors would be created by values or categories of values of the determining variables. These could randomly be selected to donate imputed income values for SHS *other adults* holding matching values for the determining variables. This method requires the values of the determining variables to be known for SHS *other adults*.

MORI consider that there are disadvantages in both the first two of these methods. For the first case, it is shown that age is major factor in the difference between HRPs' and other adults' income. Age group specific ratios though would suffer from being derived from small cell sizes and thus become less robust. For the second method, it is felt that the number of random person cases who are actually other adults is too limited, leading to too small a pool of possible donors.

MORI therefore favour the third method – a form of hot deck imputation. They also propose that to counteract occasions of limited size of some donor groups, a hierarchical level of the categorisations of the values of the determining variables be introduced – level 1 would be the tightest grouping giving smallest numbers of donors per category but highest in explanatory power, a lower level 2 which is a broader grouping giving more donors per categorisation but lower in power and a yet lower broader level 3 with further loss of power. Thus power considerations give preference for a level 1 group to be used for donation subject to enough donors available for variability under random selection. Level 2 would be used if this test failed at level 1 but was met with its own larger groupings while level 3 groups would be used if neither level 1 or 2 were acceptable. This strategy in total is described as Hierarchical Hot Deck Imputation.

1.9. Practical implementation of strategy (Sections 1 & 3 of MORI report)

1.9.1. The requirement for the implementation of the preferred strategy is the existence and comparability of determining or explanatory variables. This is considered in section 1. Here it is noted that the main non-income variables that are common between the surveys and correlated to income levels are -:

- Economic status of person;
- Age of person;
- Sex of person
- Relationship to HRP;
- Household composition;
- Tenure of housing;
- Access to car/van.

MORI were also considering the possibility of using the Scottish Household Conditions Survey instead of the SHS for this work so also identified Council Tax Banding (not on the SHS) also as a possible explanatory variable.

All these variables apart from Economic Status are directly comparable between the surveys. For the economic status variable, the main difference is that in the FRS adults who are working and in Further or Higher Education are classified as working while in the SHS they are asked to self classify themselves and are more likely to be recorded as students. FRS does also record whether adults are studying at a university or college or training for a qualification in nursing or similar. Thus FRS economic status variable can be adjusted to recode such responders as students. Table 1.2 (MORI report) shows the distribution for other adults of SHS and pre and post adjusted FRS economic statuses. After adjustment, it is clear that there is very little difference between SHS and FRS.

Thus on these set of variables, FRS and SHS are considered to be comparable.

1.9.2. A detailed analysis is made of FRS sample sizes by economic status and relationship to HRP category (Table 3.2 MORI report), mean income by Government Office Region/Scotland (Table 3.3) and mean/median incomes by relationship to HRP, economic

status, tenure and council tax band (Table 3.4). These show that relationship to HRP and economic status are two very clear determining variables for income.

1.9.3. Based on these results MORI ran a regression model for *other adults* in FRS using square root of income (to reduce positive skew) as the dependent variable and the potentially available FRS and SHS variables as explanatory variables – present as separate binary dummy variables for each category with one category as a non-appearing base (zero value for each category dummy). The result of this regression (Table 3.5 MORI report) gave an R^2 value of 0.48 indicating that 48% of the income variability was explained.

While this is a high explanatory power it was found that one variable, economic status (base as full-time employment), served to contribute the largest portion of this power. In view of this, MORI then proceeded to conduct analysis separately for each economic status.

1.9.4. The results of the separate economic status analysis of FRS *other adults* is contained in Appendix 1 of the MORI report. Because of small FRS sample sizes, the economic statuses corresponding to being non-retired economically inactive – unemployed, sort term sick, long term sick, looking after home/family and other inactive – are considered as one group.

In each case the first analysis is to examine mean income by region to deduce if there are significant regional differences from Scotland. It is found to be the case that London shows significantly higher mean incomes for all earning economic statuses – full-time and part-time employment and self-employment. This is not the case for the other economic statuses, nor does any other region show a significant difference from Scotland in any of the economic statuses.

Following this, similar regressions are conducted for each economic status group as for *all other adults* with London FRS records omitted for the three earning economic statuses. For the broad economically inactive group, the detailed economic status is added as an explanatory variable.

Results of the regressions show R^2 values between 0.10 and 0.16 except for the broad inactive group where there is a value of 0.39. In no case does being a respondent in Scotland show itself as statistically significant.

MORI propose that the results of these regressions be used to determine imputation donor group typologies (categorisations of the values of the explanatory variables) at three hierarchical levels which can be used to create groups of donors for imputation. The typology definitions they propose are described in MORI report Appendix 1.

1.9.5. On the basis of the analysis described, Ipsos MORI recommend that a hierarchical hot-deck approach for imputation from FRS of income for *other adults* in the SHS is feasible. They recommend that imputation should be carried out separately by grouped economic status – full time worker, part time worker, self employed worker, unemployed or other inactive person, retired person and student. A summary of their recommendations is given on the last page of Section 3 of their report.

1.10 ONS review and development

ONS has been contracted to review the recommendations made by Ipsos MORI, to undertake practical imputation using SAS under several years of SHS (MORI's work was in SPSS and SOLAS for imputation), to review robustness at local authority and to prepare an operational system that can be repeated each year.

This report describes the research and development phase of the work. This includes reviewing the MORI recommendation and replicating their analysis and writing SAS code to implement the procedures in experimental form. A further report Part 2 – Implementation describes an operational system giving details of the program suite and instructions for running.

2. ONS research and development

2.1 Requirement

An initial requirement in the ONS study has been to examine and quality assess the MORI developed work for *other adult* imputation and to develop it into an operational system using SAS.

Following this there is a requirement to combine the imputed values with the observed SHS records – for HRP and spouses – in order to obtain poverty estimates at local authority level together with appropriate precision measures or confidence intervals. This should form part of the operational system.

Some additional issues need to be brought into consideration.

(i) The first of these is that catering for time period adjustments needs to be examined. FRS data is made available for financial years. MORI used the 2005/06 financial year data from FRS and selected SHS data by quarters to pick a matching period. However SHS works on calendar years and it is on this basis that results are published. Therefore there is a need for this imputation to be carried out on such a period. Also SHS is considered only representative of local authorities over a two year time period so the implication of this needs to be considered.

(ii) Another timing effect is that SHS data is available considerably in advance of FRS data for the same year. It would be advantageous if imputation could be effectively carried out using earlier FRS data.

(iii) The effects of varying the implementation variables and banding needs to be further investigated. It will be possible to evaluate stability of imputations under alternative banding arrangements to understand how necessary it will be to be optimal in terms of fixing these. Decisions need to be made on a final specification of these and how they will be monitored from year to year. Stability under alternative specification may nullify requirements for year to year change.

This could become particularly important if the availability of variables changes from year to year. This has already occurred with FRS not containing data on car availability in 2006/07. Since car availability is an imputation classification variable this causes a need to vary groups.

2.2 Assessment and SAS development

MORI's work has been conducted using SPSS procedures and SOLAS for actual imputation. As a first step, these procedures were reprogrammed in SAS. In the course of this development, analyses were conducted to replicate and extend some of the results presented by MORI.

2.2.1 Overall analysis

MORI's first step was to look at the distribution of the FRS dataset and to examine how income could be modelled – after adjusting for differences in definition between FRS and SHS.

In para 1.9.3. we observe that MORI obtained an R^2 value of 0.48 in conducting a regression of all *other adults* in FRS using square root of income (to reduce positive skew) as the dependent variable and the potentially available FRS and SHS variables as explanatory variables. This was for financial year 2005/06. This has been repeated by ourselves with results that confirm the MORI analysis – we found an overall adjusted R^2 value of 0.49 for the 2005/06 data. Repetition on FRS data for 2004/05 and 2006/07 confirms consistency with values of 0.49 and 0.51 respectively. Again we found that economic status alone gave adjusted R^2 values of 0.42 for 2004/05, 0.43 for 2005/06 and 0.47 for 2006/07.

We then looked at using FRS data from calendar years by selecting records by interview date from separate FRS year files. The corresponding analysis for 2006 gives an R^2 value of 0.47 for inclusion of all potential explanatory variables.

In Table 1, we present the number of FRS records for *other adults* available by economic status by year.

	Fin. Year 2004/05	Fin. Year 2005/06	Fin. Year 2006/07	Calendar Year 2006
Full time	2198	2248	1951	2249
Part time	360	402	312	393
Self employed	197	208	173	216
Students	1041	1071	848	1069
Unemployed	356	456	341	447
Retired	397	370	372	385
Looking after home/family	79	65	70	73
Long term sick/disabled	336	321	328	346
Short term sick/disabled	31	28	34	27
Other inactive	203	185	174	202
TOTAL	5198	5354	4603	5407

Table 1 – Number of *Other Adults* in FRS responses in years 2004/05 to 2006/07 and 2006

MORI suggest that in view of the driving power of economic status and given reasonable numbers of FRS observations among these, that hot deck imputation be used separately by economic status to impute income for '*other adults*' in the SHS. The exception to this are the various inactive groups – unemployed, retired, looking after home/family, long and short term sick and disabled and the other inactive group – which should be treated together because some of these groups are small.

- **On the basis of our analysis, ONS agree with the proposal to use the hierarchical hot-deck imputation suggested and to base this separately on economic status.**

How regression can be used to define donor typologies

Before moving on to discuss the recommendations for donor typologies it is worthwhile discussing how regression helps in their definition.

MORI use the results from regression analyses for each economic status to define donor typologies for imputation groups. The idea is to use regression significance to identify which variables should be present in the specification of donor groups and how banding of values is decided.

As an example, housing tenure, features as an explanatory variable on a number of occasions. It is a categorical variable with the following categories -: owned outright, mortgaged, shared ownership, rented, rent free. In specifying the regression, one of these is chosen as the base category – here ‘owned outright’ is the one chosen. Binary dummy variables represent the other categories coded ‘one’ for the category in question and ‘zero’ otherwise. Putting housing tenure into the regression involves putting all these binary variables in.

The first observation is, does the variable add any explanatory power to the regression at all? This can be judged by an obviously increasing R^2 or more formally by an F test. If not then there is no reason for it to feature in donor typologies at all.

If it does add power then, by how much compared with other explanatory variables? The answer to this question will determine how important this variable is to be among the donation typologies. If it is very important then it needs to be included at all levels of the donation hierarchy. If it is of lower importance than it can be present at level1 but might be dropped at level2 or 3.

The third question is, if included how should the categories be banded? The answer to this is judged by the significance attached to each of the binary category variables. These actually show how significantly different they are to the omitted base category.

Consider again the housing tenure variable with base category ‘owned outright’. If the only significant category was ‘renting’ then the inference would be that if tenure appears as a donor variable then it should be in the form of two bandings – ‘renting’ and ‘any other tenure’. If the significance value is only slight then tenure could appear in this way at level1 but be dropped at level2.

2.2.2. Analysis by economic status

The results of the modelling that MORI conduct by economic status give a small explanatory power – not surprising given that in overall modelling such a high proportion of the R^2 is given by the economic status variable. For full timers, they find an adjusted R^2 of 0.16 on FRS 2005/06 data. We have repeated this analysis also finding a value of 0.16 for 2005/06. Again consistency is maintained with values of 0.16 for 2006/07 though 0.21 for 2004/05.

(i) Full timers

- In the analysis of full timers, initial summary statistics of income are obtained by Government office region/Wales/Scotland/Northern Ireland. MORI's investigation showed that all regions' mean income showed no significant differences with the exception of London which was larger (£300 per week against £207 to £249 for others). This led them to propose that FRS data from London be excluded from the imputation procedures.
- We have repeated this analysis and confirm the results for 2005/06 and find that it also applies to 2006/07 (£307 per week for London against £222 to £269 for others) and 2004/05 (£291 per week for London against £185 to £242 for others). We therefore agree that London FRS data should be excluded from the imputation procedure for full timers.
- The results of the regressions on financial year 2005/06 (excluding London) for full timers by individual variable are given in Table 2. The MORI results are shown for a regression including council tax band, ONS results are shown for models including and excluding council tax band.
- This has been repeated by ONS for FRS calendar year 2006 and the results shown in Table 3 alongside the repeated MORI results for 2005/06. An additional variable, 'access to car/van' is absent from the 2006 regressions. This is because this question was dropped from the FRS survey for the 2006/07 year and is thus not available.

The results show little difference between MORI analysis and ONS analysis, between presence or absence of council tax (despite a number of the categories showing significance), between time periods or between presence or absence of 'access to car/van' variable (despite this being significant in 2005/06).

This stability in the results is heartening.

- MORI base recommendations for the hierarchy of imputation groups, on the results of this analysis – leaving aside council tax band which is not available in SHS. So for full timers, they recommend -:

Level1 : Age (6 bandings – 16-19, 20-29, 30-39, 40-49, 50-59, 60+), Sex (2 bandings), Relationship to HRP (2 bandings – Child/Parent/Other relative against unrelated), Tenure (2 bandings – Rent against other) and Access to car (2 bandings).

Level2 : Age (4 bandings – 16-24, 25-39, 40-59, 60+), Sex (2 bandings), Relationship to HRP (2 bandings – as for Level1), Access to car (2 bandings).

Level3 : Age (4 bandings – as for Level2), Sex (2 bandings).

- These groupings do capture the essence of the significance in the regressions and seem not unreasonable. Age is by far the most significant variable, followed by availability of

car and relationship to HRP. However, due to the absence of the 'access to car/van' from 2006 we have removed this variable leaving the others unchanged. The hierarchy of imputation groups for full-timers are thus -:

Level1 : Age (6 bandings – 16-19, 20-29, 30-39, 40-49, 50-59, 60+), Sex (2 bandings), Relationship to HRP (2 bandings – Child/Parent/Other relative against unrelated), Tenure (2 bandings – Rent against other).

Level2 : Age (4 bandings – 16-24, 25-39, 40-59, 60+), Sex (2 bandings), Relationship to HRP (2 bandings – as for Level1).

Level3 : Age (4 bandings – as for Level2), Sex (2 bandings).

	Apr05 – Mar06 MORI results including council tax band			Apr05 – Mar06 ONS results including council tax band			Apr05 – Mar06 ONS results excluding council tax band		
	estimate	t-value (abs)	p-value	estimate	t-value (abs)	p-value	estimate	t-value (abs)	p-value
Constant	11.9	29.8	0.00	11.9	30.2	0.00	12.13	31.5	0.00
Relation to HRP									
Parent	0.1	0.1	0.92	1.7	1.6	0.11	1.68	1.6	0.11
Other relative	0.4	1.3	0.26	0.4	0.9	0.36	0.19	0.5	0.62
Not related	1.2	4.0	0.00	1.2	4.1	0.00	1.20	4.2	0.00
Female	-0.3	3.0	0.02	-0.4	2.5	0.01	-0.37	2.4	0.02
Age									
20-24	1.9	9.5	0.00	2.0	9.8	0.00	2.1	10.0	0.00
25-29	3.0	15.0	0.00	3.2	13.6	0.00	3.3	13.7	0.00
30-34	3.5	11.7	0.00	3.5	10.9	0.00	3.6	10.9	0.00
35-39	3.5	8.8	0.00	3.8	9.8	0.00	3.8	9.8	0.00
40-44	2.7	6.8	0.00	2.7	5.8	0.00	2.6	5.7	0.00
45-49	2.8	5.6	0.00	3.0	5.6	0.00	3.3	6.0	0.00
50-54	4.3	7.2	0.00	4.6	7.6	0.00	4.6	7.5	0.00
55-59	2.6	3.3	0.00	2.1	2.6	0.00	2.2	2.8	0.00
60-64	1.7	1.7	0.10	2.8	2.4	0.00	2.8	2.4	0.00
65-69				2.1	0.9	0.37	2.4	1.0	0.32
70-74				-1.9	0.9	0.38	-1.5	0.7	0.51
75-79				-5.3	1.6	0.12	-5.3	1.5	0.13
80-84				.	.	.	0	.	.
85+				.	.	.	0	.	.
Scottish survey	-0.3	1.5	0.93	-0.2	1.0	0.30	-0.1	0.4	0.71
Council tax band							xxxxx	xxxxx	xxxxx
B	0.3	1.5	0.17	0.3	1.2	0.23	xxxxx	xxxxx	xxxxx
C	0.7	3.5	0.00	0.7	2.9	0.00	xxxxx	xxxxx	xxxxx
D	0.8	4.0	0.00	0.7	3.1	0.00	xxxxx	xxxxx	xxxxx
E	0.5	1.7	0.06	0.5	1.8	0.07	xxxxx	xxxxx	xxxxx
F	1.5	5.0	0.00	1.6	4.2	0.00	xxxxx	xxxxx	xxxxx
G	0.6	1.2	0.24	0.5	1.0	0.34	xxxxx	xxxxx	xxxxx
H	5.5	4.6	0.00	5.8	3.9	0.00	xxxxx	xxxxx	xxxxx
Tenure									
mortgage	-0.2	1.0	0.09	-0.1	0.7	0.49	-0.2	1.0	0.32
Shared owner	0	.	.	0	.	.	0	.	.
Rent	-0.6	3.0	0.00	-0.6	2.6	0.01	-0.8	3.6	0.00
Rent free	0.5	0.7	0.57	-0.2	0.2	0.81	-0.2	0.3	0.80
Household has car/van	1.0	5.0	0.00	1.0	3.8	0.00	1.1	4.6	0.00
Adjusted R²	0.16			0.18			0.16		
Base categories	Child of HRP, male, aged 16-19, non-Scottish, not having a car/van, owning property outright, and being in council tax band A (where relevant)								

Table 2 – Comparison of regressions on root income for FRS full time *Other Adults* in 2005/06. (Shading for non-significant coefficients)

	Apr05 – Mar06 MORI results including council tax band			Jan – Dec 06 ONS results including council tax band			Jan – Dec 06 ONS results excluding council tax band		
	estima te	t- value (abs)	p- value	estima te	t- value (abs)	p- value	estima te	t- value (abs)	p- value
Constant	11.9	29.8	0.00	12.1	47.5	0.00	12.6	56.0	0.00
Relation to HRP									
Parent	0.1	0.1	0.92	-1.4	1.5	0.13	-1.3	1.4	0.17
Other relative	0.4	1.3	0.26	0.4	1.3	0.19	0.3	0.9	0.38
Not related	1.2	4.0	0.00	1.2	4.3	0.00	1.2	4.5	0.00
Female	-0.3	3.0	0.02	-0.4	2.4	0.02	-0.3	2.3	0.02
Age									
20-24	1.9	9.5	0.00	2.0	10.2	0.00	2.1	10.4	0.00
25-29	3.0	15.0	0.00	3.3	14.1	0.00	3.3	14.1	0.00
30-34	3.5	11.7	0.00	3.8	12.0	0.00	3.8	12.0	0.00
35-39	3.5	8.8	0.00	3.7	9.7	0.00	3.7	9.5	0.00
40-44	2.7	6.8	0.00	3.3	7.6	0.00	3.3	7.6	0.00
45-49	2.8	5.6	0.00	2.9	5.7	0.00	3.0	5.8	0.00
50-54	4.3	7.2	0.00	4.1	6.8	0.00	4.2	6.8	0.00
55-59	2.6	3.3	0.00	2.6	3.3	0.00	2.7	3.3	0.00
60-64	1.7	1.7	0.10	4.0	3.5	0.00	3.9	3.3	0.00
65-69				3.8	1.6	0.10	4.0	1.7	0.09
70-74				0.6	0.4	0.71	0.9	0.5	0.62
75-79				-0.1	0.0	0.98	0.5	0.2	0.83
80-84				0	.	.	0	.	.
85+				0	.	.	0	.	.
Scottish survey	-0.3	1.5	0.93	-0.3	-1.4	0.16	-0.2	0.8	0.42
Council tax band									
B	0.3	1.5	0.17	0.2	0.8	0.40	xxxxx	xxxxx	xxxxx
C	0.7	3.5	0.00	0.7	3.1	0.00	xxxxx	xxxxx	xxxxx
D	0.8	4.0	0.00	0.7	2.9	0.00	xxxxx	xxxxx	xxxxx
E	0.5	1.7	0.06	0.8	2.8	0.01	xxxxx	xxxxx	xxxxx
F	1.5	5.0	0.00	1.5	4.2	0.00	xxxxx	xxxxx	xxxxx
G	0.6	1.2	0.24	0.8	1.7	0.10	xxxxx	xxxxx	xxxxx
H	5.5	4.6	0.00	7.8	4.9	0.00	xxxxx	xxxxx	xxxxx
Tenure									
mortgage	-0.2	0.2	0.35	0.0	0.1	0.89	0.0	0.2	0.83
Shared owner	0	.	.	0	.	.	0	.	.
Rent	-0.6	3.0	0.00	-0.7	3.2	0.00	-1.0	4.5	0.00
Rent free	0.5	0.7	0.57	0.5	0.6	0.56	0.3	0.4	0.73
Household has car/van	1.0	5.0	0.00	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx	xxxxx
Adjusted R²	0.16			0.18			0.16		
Base categories	Child of HRP, male, aged 16-19, non-Scottish, not having a car/van, owning property outright, and being in council tax band A (where relevant)								

Table 3 – Comparison of regressions on root income for FRS full time *Other Adults* in 2005/06 for MORI results and 2006 for ONS results. (Shading for non-significant coefficients)

(ii) Economically inactive

- As this grouping is comprised of five different types of inactivity, there are special procedures in place. In particular, the five specific types of inactivity are considered among the determining variables.
- In the analysis of this group, there is not a significant difference in income between Government Office Regions/Wales/Scotland and Northern Ireland. In particular, London does not show a difference. Thus all FRS records can be used as potential donors.
- The results of regressions including specific types of inactivity are presented in Table 4. MORI results are for financial year 2005/06. ONS results are for calendar year 2005 – the FRS records being selected from the combination of years to achieve this.
- The regressions are not completely comparable in terms of explanatory variables – however it is heartening to see very close agreement in terms of explanatory power – adjusted R² of 39% for the MORI analysis, 37% for ONS analysis. Also both show that broad agreement in terms of significant variables – particularly that the specific economic inactivity is an important variable.
- ONS therefore agrees with the banding put forward by MORI as -:

Level 1 : Detailed economic status (5 bandings : unemployed, looking after family/home, long term sick or disabled , short term sick or disabled , other inactive), Age (6 bandings : 16-19, 20-29, 30-39, 40-49, 50-59, 60+), Relationship to HRP (2 bandings : child/parent/other relative against unrelated).

Level 2 : Detailed economic status (3 bandings : unemployed , looking after family/home , all other categories), Age (4 bandings : 16-24 , 25-39, 40-59, 60+).

Level 3 : Detailed economic status (3 bandings : unemployed , looking after family/home , all other categories), Age (2 bandings : 16-24, 25+).

(iii) Other groupings

The analysis results for other groupings is not presented here but follows in a similar manner. There are no concerns about the donation typologies. The only change made has been the necessary one to eliminate ‘access to car/van’ wherever it occurs due to its elimination from the 2006/07 FRS.

	Apr05 – Mar06 MORI results including council tax band			January – December 2005 ONS results excluding council tax band		
	estimate	t-value (abs)	p-value	estimate	t-value (abs)	p-value
Constant	2.4	4.8	0.00	2.1	5.9	0.00
Relation to HRP						
Parent	0.3	0.4	0.76	-0.4	0.6	0.57
Other relative	-0.4	1.0	0.36	0.0	0.1	0.92
Not related	1.8	4.5	0.00	1.5	3.6	0.00
Female	0.2	1.0	0.44	xxxxx	xxxxx	xxxxx
Age 20-24	1.9	6.3	0.00	1.6	5.2	0.00
25-29	2.9	7.3	0.00	2.5	5.9	0.00
30-34	3.2	6.4	0.00	3.5	7.0	0.00
35-39	5.2	8.7	0.00	4.7	7.8	0.00
40-44	3.4	5.7	0.00	3.4	5.8	0.00
45-49	4.0	5.7	0.00	3.5	5.4	0.00
50-54	5.9	8.4	0.00	5.0	7.1	0.00
55-59	4.7	6.7	0.00	4.2	5.7	0.00
60-64				5.3	6.4	0.00
65-69				7.3	4.7	0.00
70-74	5.3	6.6	0.00	0	.	.
75-79				0	.	.
80-84				0	.	.
85+				0	.	.
Scottish survey	0.8	2.0	0.05	xxxxx	xxxxx	xxxxx
Council tax band						
B	-0.9	3.0	0.00	xxxxx	xxxxx	xxxxx
C	-0.4	1.3	0.14	xxxxx	xxxxx	xxxxx
D	-0.2	0.5	0.64	xxxxx	xxxxx	xxxxx
E	-0.7	1.4	0.12	xxxxx	xxxxx	xxxxx
F	-0.8	1.3	0.17	xxxxx	xxxxx	xxxxx
G	-2.6	3.3	0.00	xxxxx	xxxxx	xxxxx
H	n.a.	n.a.	n.a.	xxxxx	xxxxx	xxxxx
Tenure						
Mortgage	-0.1	0.3	0.84	0.2	0.5	0.63
Shared owner	-2.2	0.7	0.48	-2.1	-0.6	0.58
Rent	0.5	1.7	0.08	0.9	2.9	0.00
Rent free	3.0	1.4	0.16	2.5	1.6	0.12
Type of inactivity						
Looking after home/family	2.7	5.4	0.00	2.3	4.7	0.00
Long term sick	3.4	11.3	0.00	3.4	10.7	0.00
Short term sick	0.8	1.1	0.23	0.7	0.9	0.35
Other inactive	-1.2	4.0	0.00	-1.1	3.3	0.00
Household has car/van	-0.2	0.7	0.39	xxxxx	xxxxx	xxxxx
Adjusted R²	0.39			0.37		
Base categories : Child of HRP, male, aged 16-19, non-Scottish, not having car/van, owning property outright, being unemployed and being in council tax band A (where relevant)						

Table 4 – Comparison of regressions on root income for FRS Economically Inactive *Other Adults* in 2005/06 for MORI results and 2005 for ONS results. (Shading for non-significant coefficients)

2.3 Timing effects

2.3.1 Mention has been made of the differing financial year and calendar year bases of the FRS and SHS respectively. To counter this, MORI in their investigation selected a single year of FRS (2005/06) and selected a matching period from SHS records from 2005 and 2006 data by the variable 'quarter' specifying in which quarter of the year the sample was selected.

This procedure can be switched to select a single calendar year of SHS records and selecting FRS to match by way of 'quarter' variable. In this way a calendar year of SHS imputations is made which is what is required. This has been provided for in the operational system.

2.3.2. The other timing effect to consider is the delay between availability of FRS and SHS data covering the same period. FRS can be a year behind SHS. It is possible to use earlier FRS data if the incomes on these can be suitably adjusted. An index of earnings inflation is used by DWP and is available from DWP. This is an ONS developed index based on RPI, specifically 'All items RPI excluding Council Tax and Rates (AG4111)'. It is published monthly covering the preceding twelve months. This can be applied to FRS incomes using an average of the different months' figures.

The use of this index has been tested by imputing into 2006 SHS (a) from 2006 FRS without adjustment and (b) from 2005 FRS uprated by this index. Results were not sensitive to the different sources. It is therefore the recommended procedure to use this index.

2.4 Annual evaluation

It is recommended that an annual evaluation is made for the continuing suitability of the determining variables from the FRS/SHS in forming the imputation typologies – i.e. the variables present, at each level and in which banding groups. The routines 02A_..., 02B_... and 02C_... in the operational system are provided to aid this process.

This will become a requirement if important variables currently in use in FRS or SHS are removed or redefined. This has already occurred between 2005/06 and 2006/07 FRS in the removal of the variable 'access to car' which was part of many of the MORI recommended donation typologies.

3. Implementation suite

3.1. Operational program suite

An operational suite of programs written in SAS has been supplied to perform the task of taking a year's data of SHS and to impute income estimates for the omitted *other adults*. Following imputation, a program routine calculates household estimates of equivalised income for each SHS responding household. These are then treated as direct survey estimates and an estimate of median income is determined for each local authority. Estimates of relative poverty are then calculated for each local authority on the basis of estimated proportion of households with income under 60% of calculated median. Precision estimates are determined for the estimates of relative poverty which include an element representing the additional variability caused by the imputation process.

The detailed operational description of the program suite is given in 'Part 2 – Implementation' report.

In outline, the programs are assembled as a single SAS Enterprise Guide project named *povertyests*. The program names have a numeric prefix which is the sequence in which they are run. The first is 00_Intro where initial values are set up and other parameters for a particular operational run. Here will need to be specified the year of the data for both SHS and FRS.

3.2. Time periods covered

The programs are written to estimate for calendar year time periods. Ideally both SHS and FRS will cover the same period. However the facility is provided for an earlier period for FRS to be used and an inflation figure to be supplied which can be applied to all FRS incomes for donation. The index that has been tested is the 'All Items RPI excluding Council Tax and Rates (AG4111)'. Use of this index has been found to result in only marginal changes to the final local authority poverty estimates. If used it needs to be specified in the 00_Intro program modele.

3.3. Assembling donors and recipients

The next program modules, 01_FRSPrepare to 04_AssmlAllEcostatFRSDonors are concerned with preparing and assembling the FRS donors into the donor groups at each of the three hierarchical levels for each economic status. The modules 02A_..., 02B_... and 02C_... contain data investigation and regression routines. Here the analyses which have been described to determine the donor typologies are carried out. It is not expected that there will be changes in importance of variables, however the outputs should be checked to ensure stability is maintained. Also account will need to be taken when variables change – such as the removal of the 'access to car/van' variable from the FRS in 2006/07.

After assembling the FRS donors, program 05_Create_SHS_RecipientFiles is run to assemble all the SHS records requiring income imputation. The imputation is carried out in the following program 06_IMLImputeReplicates.

3.4. The imputation process

The imputation process follows the procedures described earlier. A donor record is randomly found from the set of donor records of the same economic status and matching the recipient in the grouping variables at one of the hierarchical levels. The level selected is dependent on the number of donors available at this classification in order to ensure the random selection

gives suitable variability. The minimum number of donors available has been chosen to be ten at levels 1 and 2. At level 3 there is no minimum size specified to ensure that all recipients receive an imputed income at least at this level.

The program carries out imputation ten times for each recipient, thus ten replicates of imputed incomes are available for *other adult* recipient. The first of these is chosen as the income to be used, with the remainder used to estimate variability.

3.5. Poverty rate calculation

The final program module 07_Createpovertyestimates is where the imputed values are combined with the SHS responses from HRPs and spouses.

The procedure adopted is to create ten replicated incomes for each sampled household. For households with no *other adults* this is simply ten replications of the household income on the SHS record. For households with *other adults* whose income has been imputed, it is the ten sums of the income reported for the interviewed adults (HRP and spouse) plus each of the ten replications of imputed incomes. These incomes are then equivalised using the OECD equivalisation factors for each household.

Survey weights are then used to determine the estimated median equivalised household income for each local authority. On this basis households whose equivalised income is less than this figure are deemed to be in poverty. Survey weights are then used to determine the proportion of households in poverty for each local authority.

While these calculations are made for all ten replications, it is replication 1 which is used as the estimate.

The same procedure is used to determine the proportion in poverty in Scotland overall.

Precision measure of local authority estimates

Calculation of a precision measure for estimates of poverty is complicated by the fact that the reference value, the median of the income distribution, is itself an estimate from the survey whose standard error is not easily determined. However, it is an all-Scotland estimate and will thus have a high level of precision. For these purposes, it is taken as a fixed value. The precision measure of the local authority estimate is calculated from two components in the following way -:

The binomial distribution standard error corrected by the SHS design factor is used as one component of the precision measure i.e.

$$\text{Standard error1} = \text{Design factor} \times \sqrt{p1 \times (1-p1)/\text{household sample size}}$$

where $p1$ = estimated proportion in poverty from replication 1

The other component is determined from the variability over the ten replications

$$\text{Standard error2} = \sqrt{\text{Variance}(p1 \text{ to } p10)}$$

where $p1$ to $p10$ are the estimated proportions in poverty from replications 1 to 10 respectively.

The combined standard error is formed in the following manner

$$\text{Combined standard error} = \sqrt{\text{Standard error}_1^2 + \text{Standard error}_2^2}$$

These are then applied to the point estimates of poverty, p_1 , in the normal manner to give a confidence interval :

$(p_1 - 1.96 \times \text{Combined standard error})$ to $(p_1 + 1.96 \times \text{Combined standard error})$

4. Conclusions and recommendations

ONS endorses the estimation procedures developed by Ipsos MORI.

ONS have developed these to produce single year estimates and confidence intervals at local authority level.

The quality level of the estimates is in the most part drawn from the direct estimate portion, with the imputation having a relatively minor effect. In this way the quality is comparable to that of other similar proportion local authority SHS estimates.

The output estimates are best single year estimates with coefficients of variation ranging from around 6% (for the largest cities) to a little over 20% (for lower population areas). Most commonly they are between 10% and 13%. This means that on a 20% poverty rate the confidence interval will be of the order of ± 3 percentage points for the largest cities or ± 9 percentage points in low population areas. For the mid range CVs, it is about ± 6 percentage points. This uncertainty needs to be borne in mind in use of these outputs. It means that there could be quite wide differences in estimates between single years which are not statistically significant.

Taking more than one year at a time will, of course, reduce the uncertainty. It will however mean that the estimate is an average over the years, rather than representing a single year. A two year period will reduce the confidence interval by about a third, three years by about 40% and four years by about half.

The recommendation made in 2.4 to annually review should be noted.