Family Resources Survey

RELEASES 2000-01

RELEASE	CHANGES SINCE LAST RELEASE	RELEASE DATE
frs0001a	**FIRST RELEASE** User acceptance	14/09/2001
frs0001a	**FIRST RELEASE** Public release	14/12/2001
frs0001b	**SECOND RELEASE** The changes are in the following areas: **All total income Derived Variables (DVs) due to tax credit amendments.** - The definitional treatment of tax credits has been revised. **All investment/total income derived variables due to ISA.** - Interest from ISAs has been added to the two investment income DVs (ININV and NININV). **Pension Contributions.** - 5 cases had lump sum pensions contributions imputed – these have now been replaced with weeklyised amounts. **Self employed cases.** - About 20-25 self employed cases have been adjusted due to an error in credibility checking. **Grossing factors.** - Revised grossing factors.	21/03/2002
frs0001c	**Grossing factors on Adult and Benefit Unit Tables.** - Correction made to the grossing factors held on the Adult and Benefit Unit tables following an error in the update for release frs0001b.	31/04/2002
frs0001d	**Interim Grossing factor added** - See paper for more details	25/11/02
frs0001e	Misleading Deprivation Band Indicator (DEPBAND) removed for non-English Local Authorities. See 2002-03 Changes documentation for full details. Amendment of travel to work costs to use dual mileage rates (TTWCOST). See	24/11/03

	2002-03 Changes documentation for full details. Correction made to usual Gross pay to stop double counting deductions. (UGRSPAY). See 2002-03 Changes documentation for full details. Undo change made to INIRBEN and INOTHBEN in 2000-01 to removed benefit types 26 and 51 (Back to Work Bonus and child maintenance bonus) and instead include within benefit income and weeklyise the lump sums. This amendment is in line with HBAI treatment In 1996-97 the derived variable for specific household tenure types (TENTYPE) was changed from ten categories to eight. The format attached to this variable was not updated. This has now been corrected for all affected years. Family Type (FAMTHBAI) definition adjusted to be in line with HBAI definition introduced in 2001-02. See 2002-03 Changes documentation for full details.	
frs0001f	New Grossing regime (GROSS3) introduced - See paper for more details One case has JSA(IB) reinstated to account for deductions from benefit DV ININV – Income from Investments – has been corrected for where the code was excluding income from GILT Edged Stock when income reported after tax. Introduction of HRPID to the flatfile	22/11/04
Frs0001g	Revised weights issued for the new Grossing regime (GROSS3).	27/01/05
Frs0001h	Revised weights issued for the new Grossing regime (GROSS3) to correct for overestimation of the Lone Parent population control.	09/02/05

FAMILY RESOURCES SURVEY 2000-01

SUMMARY OF EDITING AND IMPUTATION PROCEDURES CARRIED OUT BY DWP

For the 2000-01data set, the following tasks were carried out by DWP.

1 Conversion of monetary amounts to weekly values

Many of the questions on the FRS ask for amounts received/paid and to what period they relate (e.g. benefit receipt, council tax payments). In these cases, amounts were converted to weekly equivalents.

More information on which period code relates to which value is given in the Excel spreadsheet period.xls.

1.1   During the conversion process amounts were not converted where:

     1.1.1   payments were one off or lump sum payments (period code 95)
     1.1.2   "none of the above" (period code 97)
     1.1.3   period code missing
     1.1.4   payments were less than 1 week (period code 90)

1.2   However, for those items of income and expenditure which feed in to derived variables used by the DWP, missing, 90, 95 and 97 period code payments were scrutinised and edited to a weekly value. Remaining 90, 95 and 97 period codes will appear in analyses as outliers. Users will need to consider whether to edit or delete these cases. The easiest way to identify such variables is to consult minmaxan.xls and search on maximum values of 95 or 97. The link between period codes and monetary amounts is given in period code meta.xls.

1.3   Note that in the dataset period codes shown as –1 (skipped) have an imputed weekly amount attached.

2 Validation, editing and imputation

Information about procedures carried out by DWP are contained in the Methodology chapter of the latest FRS publication.

3 Anonymisation

1.2   ONS/National Centre for Social Research have their own procedures to ensure the confidentiality of respondents. Names and addresses are kept separately from the data and are not supplied to the DWP.

1.3   Additional steps have been taken by the DWP prior to release of the data outside the department. These are:

     1.3.1   The following variables have been removed from the data set:

Variable	Table
Acorn	Househol
Grossct	Househol
Lac	Househol
Nindinc	Adult
Ninearns	Adult
Nininv	Adult
Ninpenin	Adult
DOB	Adult
DOB	Child

1.3.2   Monetary amounts relating to council tax variables have been rounded to whole pounds. Variables affected are:

Variable	Description	Table
ctamt	last CT payment	househol
ctrebamt	amount of CT rebate	househol
ctredamt	amount of transitional reduction	househol
cwatamt	amount included in rent for CT water charge	househol
indinc	Derived Variable (DV) for adult income	adult
inrpinc	DV for adult RP/IS income	adult
indisben	DV for adult disability benefit income	adult
inirben	DV for adult income related benefit income	adult
innirben	DV for adult non-income related benefit income	adult
inothben	DV for adult other benefits	adult
buinc	DV for benefit unit income	benunit
burpinc	DV for benefit unit RP/IS income	benunit
budisben	DV for benefit unit disability benefit income	benunit
buirben	DV for benefit unit income related benefit income	benunit
bunirben	DV for BU non-income related benefit income	benunit
buothben	DV for BU other benefit income	benunit
hhinc	DV for household income	househol
hhrpinc	DV for HH RP/IS income	househol
hhdisben	DV for HH disability benefit income	househol
hhirben	DV for HH income related benefit income	househol
hhnirben	DV for HH non-income related benefit income	househol
hhothben	DV for HH other benefit income	househol
hbeninc	DV for HH benefit income	househol
cwatamtd	DV for council tax water charge	househol
burent	DV for BU rent	benunit
hhrent	DV for HH rent	househol
hscosthh	DV for HH housing costs	househol

1.4   However, assurances given to interviewees allow DWP to provide unanonymised data in very restricted circumstances. For more information, please contact Steve Elliot at the address given below.

ASD3E
Analytical Services Division
Department for Work and Pensions
Room 455
The Adelphi
1-11 John Adam Street
London WC2N 6HT

IMPUTATION OF MISSING VALUES IN THE 2000-01 FAMILY RESOURCES SURVEY

INTRODUCTION

Imputation is the process in which missing values in a data set are converted to non-missing values.

When a respondent answers a particular question in a survey they can state that they don't know the answer to a question, or simply refuse to give a response. Such responses are recorded and are referred to as *'missing values'*.

These values can either be left as missing, in which case you would have gaps in your data set, or replaced (*imputed*) with an estimate of the answer that the respondent would have given if they had actually answered the question.

User requirements have deemed the latter process necessary in the Family Resources Survey (FRS). The main objective of imputation is to maximise the information available to users for analysis. Furthermore, the imputation carried out simplifies the analysis for users and helps to secure the uniformity of analysis created from the FRS data sets.

It should be noted that none of the variables in the admin, benefits and care data sets are imputed and that benefit editing is carried out separately to the rest of imputation.

Methodology

Imputation on the FRS has traditionally been carried out in four different ways. A brief overview of these methods is given here:

- **Bulk edits** – converting en masse a batch of cases with missing values that satisfy a particular characteristic to an identical value. This is a very crude method of imputation and can only be used in certain circumstances. For example, for people who don't know if they are in receipt of a particular benefit, we could:

    i)   edit the answers to yes, in which case we would have to open up a record for the particular benefit and impute answers for it

    ii)   edit such answers to no – which is known as *closing down routes* and is the default principle adopted in the imputation of such *routing* variables in the FRS.

- **Hotdecks** – examining the data set for non-missing cases which have similar characteristics to that with the missing value, and substituting one of these non-missing values for the missing case at random. It is usual for the characteristics to bear some relationship to the variable to be imputed, the theory being that all cases matching the chosen characteristics will have similar values for the variable

we are concerned with. For example we could impute rent for a household by randomly selecting a non-missing value from a case with the same number of rooms, council tax band, type of landlord and region as the case in question.

- **Algorithms** – a process in which one can predict the missing value for a particular case by looking at other relevant characteristics and applying a pre determined set of rules (e.g. modelling council tax payments based on council tax band, local authority and entitlement to discount).

- **Neural Networks** - Neural networks are information processing systems that learn by example, recognising patterns in data. Their main advantage over standard statistical techniques is that they can extract and model non-linear relationships without assuming any particular underlying distribution. This method was previously used in the FRS, but its use has been discontinued this year (see below).

Missing Values

There are four possible types of missing values in the FRS:

- **.A** – denotes a 'skipped' response. Such a reponse occurs where a respondent has not been routed to this particular question and an answer is not therefore required and imputation is not normally necessary.
- **.B** – denotes the fact that the respondent *'doesn't know'* the answer to the question and imputation will normally be required.
- **.C** – denotes a refusal to answer a question and, again, imputation is normally required.
- **.D** - is only output in the production of derived variables, and denotes either a mistake in the imputation process or faulty logic in the DV code. All .Ds in income and expenditure data are investigated and removed from the data set prior to user release.

Imputation Checking

Checks are carried to ensure that the imputation process has not changed the distribution of the data. Examples of these are as follows:

- A comparison of the means, standard deviations and minimum/maximum values for each variable is undertaken both post and prior imputation. Any large discrepancies (indicating that imputation is potentially biasing the data) are investigated.

- There can be cases in Hotdecks where we impute a large number of cases to a particular value, which is taken from one particular 'donor' case. This is a source of potential bias, and checks exist within hotdecks to monitor this. Where these checks show this to be a problem, remedial action, in the form of adjusting either the imputed value or the hotdeck, is taken.

- Finally credibility checks are run, which ensure that the data within individual cases is consistent, and feasible values have been imputed. Examples of these include:

  i) Checking that housing costs are generally less than income for cases in which components of either have been imputed.

  ii) Checking that gross income is greater than or equal to net income.

  iii) Checking that personal pension contributions are generally less than income for cases where components of either have been imputed.

Changes to Imputation 1999/2000

There have been no methodological changes for 1999-00.