Family Resources Survey

RELEASES 1998-99

RELEASE	CHANGES SINCE LAST RELEASE	RELEASE DATE
frs989a	**FIRST RELEASE**	2/03/00
frs989b	**SECOND RELEASE**. • Rounding up imputed values for some cases as should be integer. 55 cases affected, variables affected are TUBORR and GRTDIR1 on the Adult table and GRTDIR1 on the Child table. • Change to variable PAYDAT (date variable) – 24 cases changed. • Correction of DLA care benefit amount for 1 case (sernum 4698191). • Deletion of 8 Income Support records and 2 JSA(IB) records. • Minor changes to derived varables: HHRENT, FSMHH, FSMLKHH and FWMLKHH. • Other small knock on changes to benefit/total income/housing cost type derived variables due to changes mentioned above. • Addition of 3 dvs for ASD3A – TUBURENT, TUHHRENT and TUWATSEW	7/04/00
frs989c	**THIRD RELEASE**. • Definitional change to some earnings derived variables to bring them more in line with HBAI definitions. This has lead to small changes to various income derived variables. • Correction to EMPSTATI/EMPSTATC/EMPSTATB on Adult table. • Deletion of 2 IS records as two people in benefit unit receiving IS and this is not possible • Rounding up further imputed values for some cases as should be integer. 127 cases affected, variables affected are GRTVAL1 and GRTAMT1 on the Adult and Child table, ED1SUM and PAREAMT on Adult table and RMAMT on Mortgage table. • Changing of adult level benefit flag variables to make these consistent with BENEFITS table. • Correction to 3 imputed values of	5/06/00

	TTWFAR (Adult table).	
frs989d	**FOURTH RELEASE** • Correction to NINDINC / NINEARNS on Adult table. 148 cases have an increased NINDINC / NINEARNS as a result of the correction. • Other variables with minor corrections are: EMPSTATB, EMPSTATC, EMPSTATI, TAXPAYER, & TRAIN on the Adult table. DEPDEDS, ECOTYPBU, ECSTATBU, & LASTWORK on the Benunit table. NDDCTB and NDDISHC on the househol table. Continued on next page	06/12/2000
frs989e	**FIFTH RELEASE** The hierarchical data has not changed, this release is a change to the flatfile only. Payments from annuity pensions were incorrectly mapped to trust funds or missed from the flatfile. 98% (£11,267) of the payments from annuity pensions were incorrectly mapped to trust funds. The remaining 2% (£376) plus all trust fund payments (£5,161) were previously missed from the flatfile.	02/03/01
frs989f	**SIXTH RELEASE** **Interim Grossing factor added** - See paper for more details	25/11/02
frs989g	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 2002-03 Changes documentation for full details. 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.	24/11/03

	Family Type (FAMTHBAI) definition adjusted to be in line with HBAI definition introduced in 2001-02. See 2002-03 Changes documentation for full details.	
Frs9899h	New Grossing regime (GROSS3) introduced - See paper for more details	22/11/04
Frs9899i	Revised weights issued for the new Grossing regime (GROSS3).	27/01/05
Frs9899j	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 1999/00:

SUMMARY OF EDITING AND IMPUTATION PROCEDURES CARRIED OUT BY DSS

For the 1999/00 data set, the following tasks were carried out by DSS.

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 (eg 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 period35.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 DSS, 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 period34.xls.

2 **Validation, editing and imputation**

Information about procedures carried out by DSS are contained in the file 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 DSS.

1.3 Additional steps have been taken by the DSS 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 DSS to provide unanonymised data in very restricted circumstances. For more information, please contact Angela White at the address given below.

ASD3E
Analytical Services Division
Department of Social Security
4th Floor
The Adelphi
1-11 John Adam Street
London WC2N 6HT

IMPUTATION OF THE 1998-99 FAMILY RESOURCES SURVEY

1. INTRODUCTION

Imputation is the process whereby missing values, for chosen variables, are edited to valid values. Missing values occur in survey data when a respondent refuses to or does not know the answer to a particular question. 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 datasets. Variables in the admin and care datasets are not imputed. Benefit edits are carried out separately to the rest of imputation.

A combination of methods were used to impute values for missing data in the 1998-99 FRS dataset. The sections below provide an outline of the imputation process as well detailed descriptions of each of the different methods used. This document, intended for user information, applies to the imputation of missing data in all the FRS tables except BENEFITS, for which a different approach was used.

2. METHODOLOGY

The methods of imputation for 1998-99 were:

- **Bulk Edits**. For some missing values the imputed value is set to the same value in every case. This type of imputation is carried out on categorical variables, in particular where the question can open up a route to further questions. For example the variable ACCOUNT 'do you have any accounts' would be set to 'no' so that no further imputation would be required on subsequent questions about accounts.

- **Algorithms**. The algorithm may simply just set all missing values to a set value, e.g. the mean, or it may be more complex and involve formulae relating to a number of other variables, where a relationship exists.

- **Hot Decking**. Hot deck imputation is fundamentally a process of assigning a non-missing value, taken from a 'donor' case, to a 'target' case, which had a missing value for the variable of interest. The donor case has to fit the same 'factor' values as that of the target case. Factor values are variables, which will have an impact on the variable to be imputed.

- **Neural.** Neural Networks can impute data by finding patterns in the dataset and then applying those patterns to the data that require imputation.

- **Mop-up**. Any remaining variables will be imputed at this stage in the imputation process.

In general the imputation process was carried out in the order given above.
Benefit editing is carried out on those benefit amounts which should have a set value, for example standard Child Benefit amount in 1998-99 was £11.45. Any cases which are not standard benefit amounts or they do not know what amounts they are receiving, are edited to the correct value according to their individual circumstances.

3. CHECKING IMPUTATION

Checks are carried to ensure that the imputation process has not changed the distribution of the data. A comparison of the means, standard deviations and minimum/maximum values for each variable is undertaken. There can be cases where we impute a large number of cases to a particular value, which reduces the variability in the data. Checks are carried out to ensure that we are maintaining some randomness in the data. It is particularly important to check those variables which have a high number of imputations. Finally credibility checks are run which ensure that the data within individual cases is consistent, and feasible values have been imputed. An example of such a check is to ensure that water rates are lower than council tax amounts when the respondent has stated that council tax includes water rates.

3. VARIATION IN MISSING VALUES

Approximately 50% of all FRS variables had some missing values. Of these 57% were missing less than 1% of expected values. Table 1 lists the variables where 10% or more of expected completed values were missing, along with the method of imputation used for each. Table 2 lists the top 30 variables with the highest number of missing values. Those variables which appear in both tables, i.e. have more than 10% of values missing and have a high actual number of missing values are highlighted in the tables.

4. TABLES OF RESULTS

Table 3 provides an overall summary of imputation outlining the number of missing values initially and how many were imputed by each method. It also provides a comparison with previous years, where possible. Users will note a small fall in the total percentage of missing values. This is most likely to be due to improvements in question wording, deleting some of the questions that have a high item non-response and increasing the number of checks within the questionnaire.

There has been a continuing decrease in the number of missing values which require imputation over the last four years even though the required number of values has been steadily increasing. There is an overall increase in the percentage of missing values imputed, with 94% of missing values imputed in 1998-99.

The overall trend is an increase in the proportion of imputation undertaken by bulk edits and a decrease in the proportion undertaken by neural networks.

The drop in the number of hot decks undertaken in 1997-98 and the increase in the number of neural networks was due to a difference in the definition in mop up imputation. In 1997-98 mop up imputation was allocated to neural networks and therefore the number of imputation by this method was considerably higher than previous years and later years. In 1998-99 mop up imputation was allocated to whichever method the mop up was undertaken by and provides a much more accurate reflection of the type of imputation undertaken.

5. FURTHER INFORMATION

If you require any further information on imputation ask Julie Stanborough or any of the FRS team.

Julie Stanborough
ASD3E
☎ 28239

Table 1. Variable with 10% or more of expected values missing.

TABLE	VARIABLE	DESCRIPTION FROM LABELS 35	No. MISSING	% MISSING	METHOD OF IMPUTATION
JOB	SENIIAMT	Amount of National insurance lump sum	303	65%	ALGORITHM
RENTER	RENTFULL	Full rent before HB/rent rebate	17	53%	HOTDECK
JOB	HHA3	Amount of 3rd refund for HH expenditure	1	50%	ALGORITHM
CHILD	CHWKTST	Number of weeks had income from a Trust	3	50%	ALGORITHM
ADULT	LN1RPINT	1st loan repayment incl interest/capital	13	48%	ALGORITHM
RENTER	**WSINCAMT**	**Amount incl in rent for water/sewerage**	**731**	**45%**	**HOTDECK**
HOUSEHOL	**STRINS2**	**Insured value of the structure**	**342**	**42%**	**HOTDECK**
JOB	TAXDAMT	How much income tax deducted last time	87	40%	HOTDECK
JOB	PRBEFORE	Amount of profit before tax	116	39%	HOTDECK
JOB	SMPRATE	Higher/lower rate of SMP	9	38%	ALGORITHM
ADULT	APDAMT	Amount from absent partner paid directly	19	34%	NEURAL NETWORKS
ADULT	LN2RPINT	2nd loan repayment incl interest/capital	2	33%	ALGORITHM
JOB	SMPAMT	Amount included in last wage for SMP	24	30%	ALGORITHM
HOUSEHOL	STRAMT1	Amount: Insurance part of repayment	974	29%	HOTDECK
ADULT	APAMT	Amount received from absent partner	11	29%	NEURAL NETWORKS
HOUSEHOL	**STRAMT1**	**Amount: Insurance part of repayment**	**1103**	**27%**	**HOTDECK**
JOB	**PROFIT1**	**Amount of net profit or loss**	**614**	**25%**	**ALGORITHM THEN HOTDECK**
ASSETS	**HOWMUCHE**	**Value of asset (office edit)**	**1104**	**24%**	**LEAVE AS MISSING**
JOB	**SETAXAMT**	**Amount of tax in last 12mths (s-empl)**	**369**	**23%**	**ALGORITHM**
ADULT	TTWREC	Travel costs: passengers' contributions	33	22%	NEURAL NETWORKS
JOB	INSTYPE1	Permanent health insurance	335	21%	HOTDECK
JOB	INSTYPE2	Critical illness cover	335	21%	HOTDECK
JOB	NIDAMT	How much National Insurance was deducted	20	19%	HOTDECK
ADULT	OTAMT	Amt of extra income tax in last 12 mths	193	18%	NEURAL NETWORKS
HOUSEHOL	CHRGAMT2	Amount paid for feu duty	2	17%	NEURAL NETWORKS
JOB	SSPAMT	Amount included in last wage for SSP	29	14%	ALGORITHM
ADULT	KEEPPEN	Keep pension and pay contribs. if leave?	1366	14%	HOTDECK
ADULT	SHAREPAY	Amount of contrib. towards housing etc	93	14%	NEURAL NETWORKS
CHILD	CHAMTTST	Income received from a Trust	5	14%	HOTDECK
ACCOUNTS	**ACCINT**	**Interest received**	**11131**	**13%**	**HOTDECK**
ASSETS	**HOWMANY**	**Number of shares/bonds/units held**	**540**	**13%**	**HOTDECK**
JOB	**SEINCAMT**	**What is your income from this business**	**306**	**13%**	**HOTDECK**
JOB	UBONAMT	Amount of bonus included in usual pay	5	13%	NEURAL NETWORKS
ADULT	ACCSSAMT	Amount : Access Fund	1	13%	ALGORITHM
ADULT	MNTUSAM1	Maintenance to self - usual amount	1	13%	ALGORITHM
MAINT	MRUAMT	Amount of usual maintenance payment	1	13%	ALGORITHM
MORTGAGE	INCMSTY3	Year MPP taken out	1	13%	LEAVE AS MISSING
ASSETS	ISSVAL	Value of NSC	14	12%	HOTDECK
ADULT	NEWDTYPE	Which new deal option	4	11%	GATEKEEPER BATCH EDIT
JOB	**GRSOFAR**	**Taxable gross earnings so far this year**	**996**	**11%**	**LEAVE AS MISSING**
BENEFITS	NOTUSAMT	Amount of benefit usually received	11	11%	BENEFIT EDITING
RENTCONT	ACCAMT	Amount of contribution to rent	7	11%	HOTDECK
ADULT	ED2AMT	Amount of last repayment on second loan	1	10%	ALGORITHM
ADULT	ED2SUM	Amount originally borrowed: 2nd loan	1	10%	ALGORITHM
ADULT	GRTVAL1	Current value of 1st grant (incl fees)	7	10%	ALGORITHM
JOB	OWNAMT	Amount of drawings from business	87	10%	HOTDECK
ASSETS	ISSDATE	Date NSC issued	56	10%	LEAVE AS MISSING
HOUSEHOL	SUBRENT	Amount of rent from subletting	4	10%	NEURAL NETWORKS

Table 2. Top thirty variables with the highest number of expected values missing.

TABLE	VARIABLE	DESCRIPTION FROM LABELS 35	No. MISSING	% MISSING	METHOD OF IMPUTATION
ACCOUNTS	**ACCINT**	**Interest received**	**11131**	**13%**	**HOTDECK**
ASSETS	HOWMUCH	Value of asset (respondent's estimate)	1936	9%	HOTDECK
BENEFITS	BENAMT	Benefit amount	1416	5%	BENEFIT EDITS
ASSETS	**HOWMUCHE**	**Value of asset (office edit)**	**1104**	**24%**	**LEAVE AS MISSING**
JOB	PAYAMT	Amount of last take home pay	1008	5%	HOTDECK
HOUSEHOL	**STRAMT1**	**Amount: Insurance part of repayment**	**1003**	**27%**	**HOTDECK**
JOB	**GRSOFAR**	**Taxable gross earnings so far this year**	**996**	**11%**	**LEAVE AS MISSING**
JOB	PAYE	Amount deducted for PAYE	981	5%	HOTDECK
JOB	NATINS	Amount deducted for NI	969	5%	HOTDECK
HOUSEHOL	STRAMT2	Amount: Insurance premium	845	7%	HOTDECK
RENTER	**WSINCAMT**	**Amount incl in rent for water/sewerage**	**731**	**45%**	**HOTDECK**
JOB	**PROFIT1**	**Amount of net profit or loss**	**614**	**25%**	**ALGORITHM THEN HOTDECK**
PENSION	PENPAY	Amount of last payment from pension	611	8%	HOTDECK
HOUSEHOL	WSEWAMT	Combined water/sewer rates: amount paid	608	4%	HOTDECK
ADULT	TTWMOD1	Travels to work by walking/bicycle	597	3%	NEURAL NETWORKS
ADULT	TTWMOD2	Travels to work by car/van	597	3%	NEURAL NETWORKS
ADULT	TTWMOD3	Travels to work by motorcycle	597	3%	NEURAL NETWORKS
ADULT	TTWMOD4	Bus/train/tube	597	3%	NEURAL NETWORKS
ADULT	TTWMOD5	Works bus/company transport	597	3%	NEURAL NETWORKS
ADULT	TTWMOD6	Travels to work by other means	597	3%	NEURAL NETWORKS
ASSETS	**HOWMANY**	**Number of shares/bonds/units held**	**540**	**13%**	**HOTDECK**
MORTGAGE	MORTLEFT	Amount outstanding on mortgage	531	6%	HOTDECK
ENDOWMNT	MENPOLAM	Last premium on endowment policy	503	6%	HOTDECK
INSURANC	POLAMT	Premium: Amount Paid on this Policy	418	7%	ALGORITHM
JOB	**SETAXAMT**	**Amount of tax in last 12mths (s-empl)**	**369**	**23%**	**ALGORITHM**
HOUSEHOL	**STRINS2**	**Insured value of the structure**	**342**	**42%**	**HOTDECK**
JOB	**SEINCAMT**	**What is your income from this business**	**306**	**13%**	**HOTDECK**
PENSION	PTAMT	Amount of tax deducted at source	305	6%	HOTDECK
JOB	DEDUC1	Amount deducted:pensions/superannuation	284	4%	HOTDECK
PENSION	PENTAX	Whether tax deducted at source on PENPAY	270	3%	NEURAL NETWORKS

Table 3. Summary of imputation

	required no	Missing no	imputed by:				Total number of	Total number of values
			bulk edit	algorithm	hot deck	neural network	Values imputed	left as missing
1995/6	**11,279,900**	**57,100**			32,100	10,200	42,300	14,800
		0.51%			56.22%	17.86%	74%	26%
1996/7	**11,642,424**	**51,917**	1,883	7,275	28,189	7,490	44,837	7,080
		0.45%	3.6%	14.0%	54.3%	14.4%	86%	14%
1997/8	**12,205,394**	**48,987**	2,313	5,779	24,744	10,717	43,553	5,434
		0.40%	4.7%	13.3%	50.5%	21.9%	89%	11%
1998/9	**12,317,734**	**44,608**	**2,690**	**5,691**	**27,973**	**5,691**	**42,045**	**2,563**
		0.36%	**6.0%**	**12.8%**	**62.7%**	**12.8%**	**94%**	**6%**