2007 PSID Income and Wage Imputation Methodology

Denise Duffy

Survey Research Center - Institute for Social Research
University of Michigan

July, 2011

2007 PSID Income and Wage Imputation Methodology

Contents

1. Introduction

The Panel Study of Income Dynamics (PSID) is a national longitudinal study of families and individuals that began in 1968. The initial focus of the PSID was to examine dynamics of employment, earnings, and income over the life cycle through interviews with roughly 5,000 families. The PSID continues to interview many of these same families today, as well as their descendents. The PSID provides a wide variety of information about both families and individuals collected over the span of the study. The central focus of the data is economic and demographic, with substantial detail on income sources and amounts, employment, family composition changes, and residential location. As such, it is critical to understand the methods used for, and potential impact of imputation of income and related employment measures in the PSID. This document provides an overview of Total Family Income, as well as wage-related employment data (work hours, wages, weeks worked) reported in the 2007 PSID, as well as a detailed description of the methods used to impute missing and unreliable income and employment data in the 2007 PSID.

2. Income and Employment Data Overview

The variable "Total Family Income" is created for each wave of PSID data collection. This generated variable is the aggregation of many sources of income across members of a PSID family. At a high level, Total Family Income is the sum of the following components of income (2007 Family File codes included):

- *Head & Wife Taxable Income (ER40943)*
- *Head & Wife Transfer Income (ER40992)*
- *Taxable Income, Other Family Member (ER40999)*
- *Transfer Income, Other Family Member (ER41020)*
- *Head Social Security Income (ER41022)*
- *Wife Social Security Income (ER41024)*
- *Other Family Member Social Security Income (ER41026)*

These income aggregates have been provided to users since the study began in 1968; therefore, we continue to provide these subtotals for consistency over the decades. We recognize that some classification names do not accurately reflect their components. For example, transfer income includes pension and annuity income. However, we keep the designations for historical consistency. Many of these high-level income sources are themselves generated variables formed by aggregating more detailed income sources.

In addition to asking respondents for detailed income data by source, the employment section of the PSID (sections BC/DE) asks respondents for detailed information for each of the Head's and Wife's jobs. These employment data are combined with the income data (above) to compute a wage rate for both the Head and Wife, and are also used for imputing labor income when necessary. The PSID asks respondents for information on

weeks worked, weeks not working by time off category (illness, unemployment, etc.), hours worked, and earnings. Respondents are asked to report on the Head's and Wife's current main job, as well as three additional jobs. If a Head or Wife holds more than four jobs, they are asked to provide information on all other jobs in place of the fourth job. In 2007, key hours, weeks worked and wage variables in the employment section include the following:

- *Head Earnings (ER36176, ER36208, ER36240, ER36272)*
- *Wife Earnings (ER36434, ER36466, ER36498, ER36530)*
- *Head Hours/Week (ER36170, ER36202, ER36234, ER36266)*
- *Wife Hours/Week (ER36428, ER36460, ER36492, ER36524)*
- *Head Weeks Worked (ER36168, ER36200, ER36232, ER36264)*
- *Wife Weeks Worked (ER36426, ER36458, ER36490, ER36522)*

3. Imputation Methodologies Used for Income and Employment Data in the 2007 PSID

Because detailed income sources are imputed individually, with varying methodologies depending on the source, we will review each income source separately. First, however, we will briefly review the methodologies used to impute income and employment variables in the 2007 PSID Family file. When a component of the generated variable Total Family Income has a value of Don't Know, NA; Refused, or is outside of the range of plausible values, the component is imputed in one of a number of ways. Because of differences in the availability of supplemental data, the number and variability of observations with non-missing data, and for other processing reasons, the granular income components are imputed using varying methodologies, and more than one imputation strategy may be used for a given component. The methods employed to impute the various components of Total Family Income include the following:

1. **Alternate Variable:** Use other data collected in the PSID to calculate the income variable of interest. In particular, data collected on wage and salary income for individual jobs is used to calculate total wage and salary income when the question asking about total wage and salary income across all jobs requires imputation (accuracy code=XSCID)
2. **Prior Wave Carryover:** Use an individual's non-missing income from the prior wave (adjusted by inflation) to impute current wave missing income (accuracy code=XYRID)
3. **Group Mean Replacement:** Impute missing income based on average income of a similar, non-missing set of observations. This is typically implemented in the PSID income data using the mean of non-missing observations from the same occupation or industry (accuracy code=ASGND)
4. **Hotdeck Replacement:** Select a non-missing observation at random from a defined set of non-missing observations (accuracy code=HTDCK)
5. **Median Replacement:** Use the overall median of non-missing observations to impute missing observations (accuracy code=MEDN)

Each of the five methods above is used to impute one or more of the components of Total Family Income. This document details the method(s) used to impute each component of Total Family Income, as well as the prevalence of imputed income by each component.

4. Head and Wife Taxable Income

Head and Wife Taxable Income is comprised of three main sources of income: Head/Wife Earnings, Head/Wife Income from Assets, and Net Profit from Farm or Business. In addition, these three main sources of taxable income have their own sub-components of income. Table 1 below details the lowest level of disaggregation of the taxable income hierarchy.

Table 1: Components of Head and Wife Taxable Income

Head Wage and Salary Income	Wife Wage and Salary Income	Head Income From Assets	Wife Income From Assets	Net Profit from Farm or Business
Head Wages (ER40903)	Wife Wages (ER40933)	Head Interest Income (ER40926)	Wife Interest Income (ER40939)	Head Net Business Income (ER40900,ER40901)
Head Bonus (ER40905)		Head Dividend Income (ER40924)	Wife Dividend Income (ER40937)	Wife Net Business Income (ER40930,ER40931)
Head Overtime (ER40907)		Head Rental Income (ER40922)	Wife Rental Income (ER40935)	Net Income from Farm (ER40898)
Head Tips (ER40909)		Head Trust Funds (ER40928)	Wife Trust Funds (ER40941)	
Head Commissions (ER40911)				
Head Market Gardening (ER40915)				
Head Professional Practice (ER40913)				
Head Roomers and Boarders				

4.1　Head/Wife Earnings

4.1.1.　Head/Wife Wage and Salary Income

In section G of the PSID Family file, respondents are asked to report wage and salary income for both the Head and the Wife by asking "How much did you [HEAD/WIFE] earn altogether from wages or salaries in 2006, that is, before anything was deducted for taxes or other things?" In 2007, the variables in the 2007 PSID (prior to imputation) are ER36928/ER37299. The methodology used to impute wage and salary income is the same for both Heads and Wives. Wage and salary income is imputed under the following conditions:

1. Wages and salaries top coded at $9,999,997 (In 2007, there were no top-coded observations)
2. Don't Know (9,999,998)
3. Refused (9,999,999)
4. Value < $0

When wage and salary income for either a Head or Wife is set to one of the values above by the interviewer, wage and salary income is imputed according to the following methodology:

Step 1: Calculate wage and salary income using wage rate information collected on individual jobs in the Employment section (sections BC and DE) of the 2007 Family file. The Employment section of the PSID asks respondents for wages, weeks worked, hours worked and time off detail for up to four jobs held by the Head (section BC) and Wife (section DE). If Employment section data are not available, then go to Step 2.

Step 2: Impute using the wage rate from the prior wave, adjusted by the change in the average annual CPI (CPI-U) from the prior wave. If not available, then go to Step 3.

Step 3: Apply overall Head/Wife median wage and salary income of the non-zero, non-imputed cases to remaining missing observations.

Table 2 and Table 3 on the following pages detail the process used for wage imputation, as well as the frequency of imputation for 2007 Heads and Wives.

Table 2: Head of Household Wage and Salary Income

Imputation Process in the 2007 PSID

```
                    ┌─────────────────────┐
                    │  Head of Household  │
                    │        8289         │
                    └──────────┬──────────┘
             ┌─────────────────┴─────────────────┐
┌────────────────────────────┐      ┌────────────────────────────┐
│ Has Wages (g13), but wages │      │        Valid Wages         │
│ set to $9,999,998 (DK,     │      │           7745             │
│ N=386) or $9,999,999       │      └────────────────────────────┘
│ (Refused, N=158)           │
│            544             │
└─────────────┬──────────────┘
       ┌──────┴───────────────────┐
┌──────────────────┐   ┌────────────────────────────┐
│ Valid 2007 BC    │   │ No Valid 2007 BC Hours and │
│ Wage and Hours   │   │         Wage Data          │
│ for Imputation   │   │            340             │
│       204        │   └─────────────┬──────────────┘
└──────────────────┘          ┌──────┴──────────────┐
                    ┌──────────────────┐   ┌────────────────────────┐
                    │ Has Valid 2005   │   │ Does not have Valid    │
                    │ Wage data        │   │ 2005 Wage Data         │
                    │      126         │   │          214           │
                    └──────────────────┘   └───────────┬────────────┘
                                           ┌────────────────────────┐
                                           │ Use Head Median Wage   │
                                           │          214           │
                                           └────────────────────────┘
```
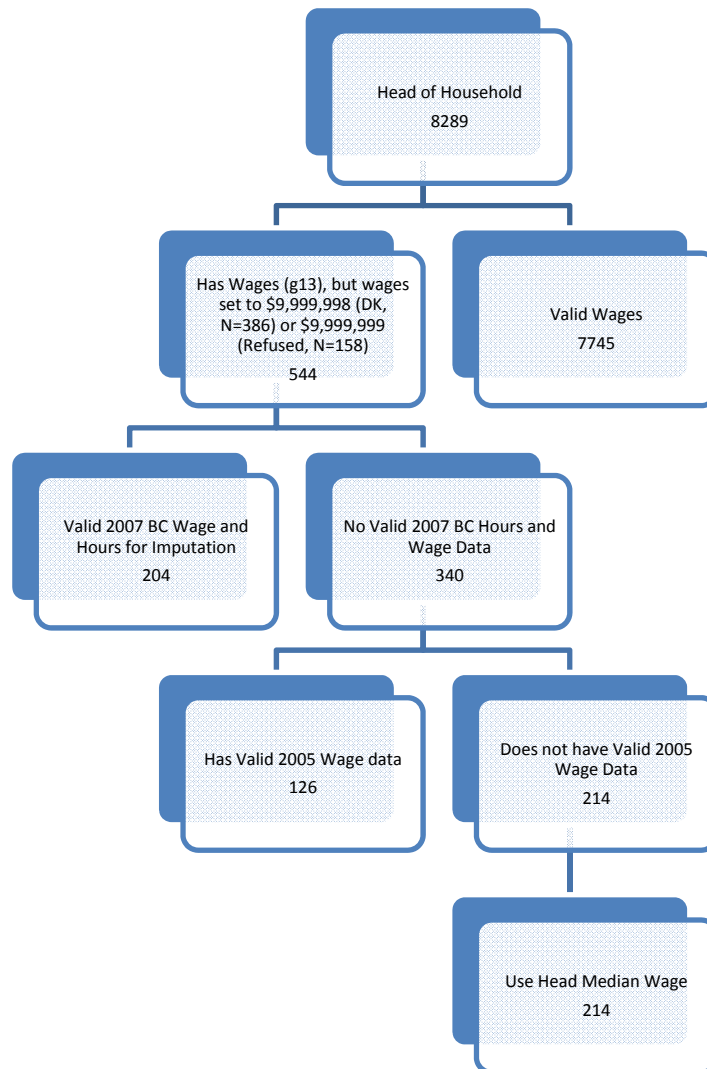
Table 3: Wife Wage and Salary Income Imputation Process in the 2007 PSID

```
                          ┌─────────────────┐
                          │      Wife       │
                          │      4548       │
                          └────────┬────────┘
              ┌────────────────────┴────────────────────┐
   ┌──────────────────────────┐              ┌──────────────────────┐
   │ Has Wages (g13), but     │              │     Valid Wages      │
   │ wages set to $9,999,998  │              │ 4298 (3101 with      │
   │ (DK, N=181) or $9,999,999│              │ wages>0)             │
   │ (Refused, N=69)          │              └──────────────────────┘
   │          250             │
   └────────────┬─────────────┘
        ┌───────┴────────────────┐
┌──────────────────┐   ┌──────────────────────┐
│ Valid 2007 DE    │   │ No Valid 2007 DE     │
│ Wage and Hours   │   │ Hours and Wage Data  │
│ for Imputation   │   │        172           │
│       78         │   └──────────┬───────────┘
└──────────────────┘       ┌──────┴───────────────┐
                 ┌──────────────────┐   ┌──────────────────────┐
                 │ Has Valid 2005   │   │ Does not have Valid   │
                 │ Wage data        │   │ 2005 Wage Data        │
                 │       52         │   │        120            │
                 └──────────────────┘   └──────────┬────────────┘
                                           ┌──────────────────────┐
                                           │ Use Wife Median Wage │
                                           │        120           │
                                           └──────────────────────┘
```
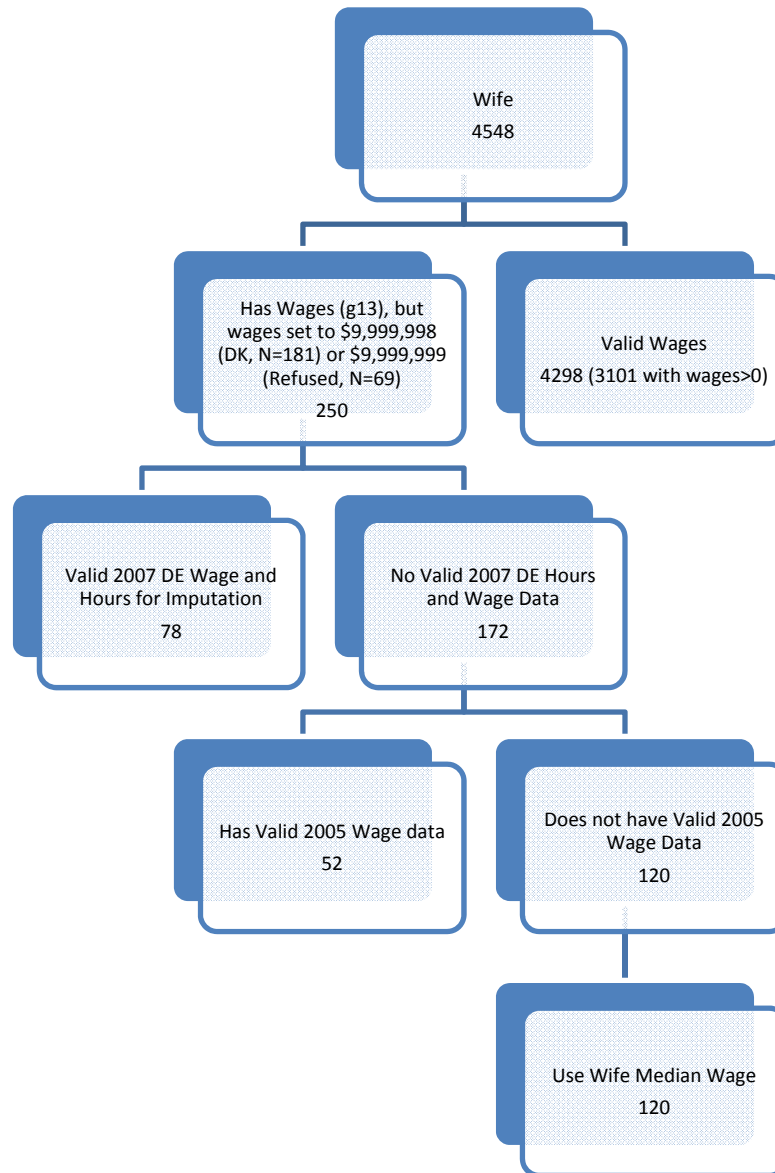
As noted above, the PSID also collects wage and salary income information by job in the employment section of the PSID, and that data is used to impute total wage and salary income when total wage and salary income is missing. In some cases, the Employment section data on individual jobs that are used to impute Head and Wife wage and salary income in Step 1 above also require imputation. In Section 9, we detail the methodology used to impute the weeks worked (including time off from work), weeks unemployed or out of the labor force, hours worked and wage rate variables from the Employment sections BC and DE of the 2007 PSID.

4.1.2. Overtime, Tips & Commission Imputation

For Heads only, additional sources of earnings including overtime (ER40907), tips (ER40909), commissions (ER40911), bonuses (ER40905), market gardening (ER40915), professional practice (ER40913), and roomers and boarders (ER36982) are imputed. (PSID is considering asking respondents for income from these sources for wives in future waves.) The imputation methods applied to overtime, tips and commissions are identical and so the sources are presented together. For each component, income is imputed under the following conditions:

1. Income >$999,996
2. Income < $0
3. Don't Know (999,998)
4. Refused (999,999)

When the overtime, tips or commissions fall into one of the above values, the income component is imputed according to the following methodologies:

Step 1: Impute using data on overtime, tips and commissions earned from the individual jobs data reported in the Employment section (section BC). If overtime, tips or commissions are not available from the Employment section, impute using Step 2.

Step 2: Calculate ratio of income to wages by income source for each individual with non-missing, non-zero income. Impute using the mean ratio by 3-digit occupation code. (see ER36132). If average income by occupation code is not available for the income source and occupation, impute using Step 3.

Step 3: Impute using overall median income amount.

Table 4 below details the income imputation process and prevalence for overtime, tips and commission income.

Table 4: Overtime, Tips and Commission Imputation

Income Source	# Heads with non-zero income	Impute Using Individual Jobs Data (Step 1)	Impute Using Average Income as a % of Wages by Occupation (Step 2)	Impute Using Overall Median Income (Step 3)
Overtime	201	21	18	21
Tips	71	1	13	0
Commission	70	1	0	13

4.1.3. Bonus Income

Head bonus income (ER40905) is imputed when respondents report that bonus income>$999,996, bonus income<$0, or the respondent answers Don't Know (999,998) or Refused (999,999). Bonus income is imputed first using the average bonus percent by occupation code (Step 1), and if average bonus percent by

occupation isn't available, using an overall median bonus amount (Step 2). Table 5 below details the frequency of Head bonus imputation.

Table 5: Head Bonus Imputation

Heads with non-zero Bonus income	585
Heads requiring Bonus income imputation	32
Use average bonus % by occupation (Step 1)	26
Use overall median bonus amount (Step 2)	6

4.1.4. Market Gardening, Professional Practice, Roomers and Boarders

The next three sub-components of Head earnings include Professional Practice, Market Gardening and Roomers and Boarders, and they are fairly small in terms of prevalence among Heads. Among the 8,289 Heads in 2007, only 92 had Professional Practice Income, 34 had Market Gardening Income, and none had Roomers and Boarders income. Of the 126 cases of non-zero income from these sources, 8 require imputation. We are able to use prior wave income for one observation (Step 1), and apply the overall mean hourly rate by income source for the remaining 7 observations (Step 2).

**Table 6: Professional Practice, Market Gardening and Roomers –
Boarders Imputation**

	Imputation Condition	#Heads with non-zero income in 2007	Step 1: Use Prior Wave Income for Head	Step 2: Use Mean hourly rate. If hours not available, use 500 (done for 1 observation)
Professional Practice (ER40913)	Income outside the range of $0 and $9,999,996, or Don't Know (9,999,998) /Refused (9,999,999)	92	1	4
Market Gardening (ER40915)	Income outside the range of -$50,000 and $9,999,996, or Don't Know (9,999,998; -999,998) /Refused (9,999,999;-999,999)	34	0	3
Roomers and Boarders (ER36982)	Income outside the range of $0 and $9,999,996, or Don't Know (9,999,998) /Refused (9,999,999)	0	0	0

4.2 Head/Wife Income From Assets

Respondents are asked to report about income from the following income sources for both Heads and wives:
- Interest (ER40926/ER40939)
- Dividends (ER40924/ER40937)
- Trust Funds (ER40928/ER40941)
- Rent (ER40922/ER40935)

If the respondent reports that the income from these sources is earned jointly by the Head and the Wife, then the Wife is allocated 50% of the income and the Head is allocated 50% of the income. If the respondent reports that the Wife receives additional income beyond that reported by the Head, then this amount is added to the joint income. Income from assets is imputed under the following conditions:
1. Income<$0
2. Don't Know (999,998)
3. Refused (999,999)

All sources of income from assets: interest, dividends, trust funds and rent, are imputed using the overall median amount of income by source for Heads and Wives. The table below details the number of imputations made by each asset source. As the table indicates, interest income is the most prevalent source of asset income, while dividend and interest income together account for nearly 97% of imputed observations.

Table 7: Asset Income Imputation

	# Head/Wifes with Income>0	# Heads/Wifes with Imputed Values
Rent Head	486	13
Dividends Head	1211	182
Interest Head	2374	326
Trust Head	81	8
Rent Wife	182	6
Dividends Wife	641	95
Interest Wife	1325	165
Trust Wife	13	0

4.3 Net Income from Farm or Business (For Head, Wife & Other Family Members)

4.3.1. Net Income from Business

For each of up to five businesses, we ask respondents to report business gross receipts (ER36867, ER36881, ER36895, ER36909, ER36923), expenses (ER36868, ER36882, ER36896, ER36910, ER36924), and profits/losses (ER36869, ER36883, ER36897, ER36911, ER36925). Specific questions in the 2007 PSID include the following:

- G11. What were the total receipts from the business in 2006?
- G11a. What were the total operating expenses, not counting living expenses?
- G11b. That left a net income of?

Net business income is viewed from the perspective of the family. A business can have a number of different ownership structures: Head owns, Wife owns, Head and Wife owns, Head, Wife and Other Family Members (OFUM) owns, etc. The process of income imputation for Net business income involves two steps: 1: impute income and 2: allocate income among family members and between labor and asset income. Table 8 below summarizes the possible net business income

ownership levels in the 2007 data, along with the number of businesses requiring imputation. Net business income is imputed when Net Income from Business is reported as follows:

1. Don't Know (-999,998; 9,999,998)
2. Refused (-999,999; 9,999,999)

Table 8: Net Business Income Imputation Prevalence by Family Business Ownership Type

Ownership	# Businesses with Non-missing Net Income	# Businesses with Imputation Required
Head	337	85
Wife	110	22
Head & Wife	74	25
OFUM Only	6	7
Head & OFUM	2	0
Head, N/A	7	6
Head, Wife, N/A	2	0
Head, OFUM, N/A	0	1
Mystery - All N/A	1	0

Net business income is imputed according to the following steps:

Step 1: Impute net business income using self employment income from detailed jobs data in the Employment section (when only one job is listed for self employment).

Step 2: If it is not possible to use the Employment section, and if the family's last wave's business was in the same industry as the current wave's business, then impute net business income using prior wave's income (adjusted by change in CPI_U from prior wave).

Step 3: If prior wave's data not available, impute net business income using the Hot Deck approach within 2-digit Census industry code (see 2007 PSID variable ER36857).

A summary of cases requiring imputation and the method used is provided in Table 9 below:

Table 9: Net Business Income Imputation Methodology

Methdology	Number of Businesses Imputed
Using Self Employment data from Jobs section (Step 1)	13
Use Prior Wave's Net Business Income (Step 2)	5
Hot Deck Methodology (Step 3)	128

The Hot Deck methodology for income imputation assigns a random donor from a set of non-missing observations to the missing observation. In the case of net business income, businesses with missing net income are randomly assigned income within an industry code. When a business reports a loss of more than -$999,997, or if the respondent indicates that the business lost money, but does not know the amount, or refuses to provide the amount, the random donor business is selected among businesses within the industry that reported a net loss. When net income is reported as Don't Know or Refused (but positive), the donor business is selected among businesses within the industry that reported a net profit. A summary of the Hot Deck approach, and the number of cases by missing status, is shown in Table 10 below:

Table 10: Net Business Income Hot Deck Imputation Methodology

Cases for Which we Impute	Hot Deck Method	# Cases
Don't Know Loss (-999,998)	Assign Random Negative Income	7
Don't Know Gain (9,999,998)	Assign Random Positive Income	106
N/A, Refused Gain(9,999,999)	Assign Random Positive Income	15

Once net business income is imputed, it is then split between labor and capital income, and allocated to Heads/Wives/Other Family Members. First, net business income is split equally among Head/Wife and Other Family Member owners. Next, income is divided between labor and asset income for Heads and Wives only according to the following rules:
- If Head or Wife responds that he or she works in the business, then his or her share of income is split evenly between labor and asset income.
- If Head or Wife does not work in the business then his or her share of income is assigned to asset income
- If the business reports a net loss, then 100% of the loss is assigned to asset income.

For other family members, net business income is split evenly between labor and asset income, unless the business realized a loss. In which case, the net loss is allocated to asset income.

4.3.2. Net Profit from Farm

Net profit from farm (ER40898) is imputed in a multi-step manner. The number of Heads with farm income is rather small (35), and the number of Heads with farm income requiring imputation is smaller still (6). The method used for imputing farm income is a follows:

Step 1:   Impute using farm receipts less farm expenditures

Step 2:   If farm receipts and farm expenditures are not available, impute using prior wave farm income adjusted for inflation using the change in CPI_U.

Step 3:   If prior wave farm income is not available, impute using the overall median farm income based on Heads with non-missing, non-zero farm income.

Of the six Heads with farm income requiring imputation, we are not able to impute any cases using receipt/expenditure data, we can impute 5 cases using prior wave farm income available, and we must impute one observation using the overall median farm net income.

5. Head and Wife Transfer Income

Transfer income is imputed using overall median substitution by income source and recipient (Head or Wife). Transfer income is imputed when a respondent reports a value of Don't Know (999,998) or Refused (999,999) for any transfer income source. A summary of the number of Head's and Wife's requiring imputation of transfer income by source is detailed in Table 11:

Table 11: Transfer Income Imputation

Transfer Income Source	Who	Number where income amount>0	Number Imputed
Alimony (ER40964)	Head	38	0
Annuity (ER40954)	Head	147	15
Child Support (ER40962)	Head	417	13
Help Non-Relatives (ER40968)	Head	443	40
Help Relatives (ER40966)	Head	962	76
Other Transfer Income (ER40970)	Head	221	18
Retirement (ER40952)	Head	772	39
SSI (ER40946)	Head	218	11
TANF (ER40944)	Head	129	2
Unemployment (ER40958)	Head	333	13
VA Pension (ER40950)	Head	211	16
Welfare (ER40948)	Head	46	4
Workers Comp (ER40960)	Head	84	5
Child Support (ER40984)	Wife	146	7
Help Non Relatives (ER40988)	Wife	12	2
Help Relatives (ER40986)	Wife	151	13
Other Transfer Income (ER40990)	Wife	138	8
Pension (ER40978)	Wife	206	8
SSI (ER40974)	Wife	43	1
TANF (ER40972)	Wife	10	1
Unemployment (ER40980)	Wife	118	5
Welfare (ER40976)	Wife	8	1
Workers Comp (ER40982)	Wife	22	1

6. Taxable Income, Other Family Member

Taxable income for Other Family Members (OFUMs) includes the following:
1. Labor Income (ER400995)
    o   Labor income from jobs held (up to four jobs)
    o   Labor income from business (imputation methodology discussed in net business income section above)
2. Asset Income (ER40997)
    a. Interest income
    b. Asset income from business (imputation methodology discussed in net business income section above)

The methodology used to impute OFUM net business income is discussed above. The overall median replacement method is used to impute labor income from jobs held, as well as interest income. Both labor and interest incomeare imputed when the following conditions hold:

1. Income <$0
2. Don't Know (999,998)
3. Refused (999,999)

Labor income is imputed for up to four OFUM jobs. Median replacement is job number specific, so, for example, imputation for OFUM's second job, would be carried out by assigning the median income across OFUM second jobs with income>$0 to that job.

Table 12: Labor Income Imputation for Other Family Members

	# Jobs with income>0	# Imputations
Job 1	2352	756
Job 2	452	156
Job 3	98	37
Job 4	29	15

For the case of asset income (interest income), 8 OFUMs have imputed interest income, while 80 OFUMs have interest income > $0.

7. Transfer Income, Other Family Member

As with OFUM taxable income, OFUM transfer income is imputed using the overall median substitution method. The frequency of imputations by source of OFUM transfer income is detailed in Table 13 below.

Table 13: Transfer Income Imputation for Other Family Members

	# OFUMS with Income Source>0	# Imputations
ADC (ER41000)	19	4
SSI (ER41001)	191	9
Welfare (ER41004)	28	3
VA Pension (ER41006)	8	4
Pension (ER41008)	61	20
Unemployment (ER41010)	10	0
Workers Comp (ER41012)	4	0
Child Support (ER41014)	28	7
Support from Relatives (ER41016)	35	3
Other (ER41018)	145	20

8. Social Security, Head, Wife and Other Family Members

Finally, Social Security Income for Heads (ER41021), Wives (ER41023) and OFUMs (ER41025) is imputed using an overall median substitution. Of the 2,168 Heads, Wives and OFUMs with positive Social Security Income, 168 require imputation.

Table 14: Social Security Imputation

	# with income>0	# Imputations
Head (ER41021)	1329	96
Wife (ER41023)	449	20
OFUM (ER41025)	390	52

9. Weeks Worked, Hours Worked and Wage Rate Imputation

This chapter details the process used for imputing the weeks worked, hours worked and wage rate variables in the Employment section (BC/DE) of the 2007 PSID. These variables are used in Step 1 of the imputation process for Head and Wife labor income (see Table 2 and Table 3).

9.1 Overview: Imputation Methods for Weeks Worked/Weeks Not Working and Hours Worked

The employment section (BC/DE) asks respondents for detailed information for each of the Head's and Wife's jobs, including information on weeks worked, weeks not working work by time off category (illness, unemployment, etc.), and hours worked. Respondents are asked to report on the Head's and Wife's current main job, as well as three additional jobs. If a Head or Wife held more than four jobs during 2006, they are asked to provide information on all other jobs in place of the fourth job. In 2007, the specific hours and weeks variables in the employment section include the following:

Table 15: Weeks and Hours Variables

Variable Name	Variable Description	Variable Group
ER40873	HEAD WORK WEEKS-2006	Weeks
ER40874	HEAD WEEKLY WORK HOURS-2006	Hours Worked
ER40875	HD OVERTIME WORK HOURS-2006	Hours Worked
ER40876	HEAD TOTAL HOURS OF WORK-2006	Hours Worked
ER40877	HD WEEKS MISSED FOR ILLNESS OF OTRS-2006	Weeks
ER40878	HD WEEKS MISSED FOR OWN ILLNESS-2006	Weeks
ER40879	HEAD WEEKS OFF FOR VACATION-2006	Weeks
ER40880	HEAD STRIKE WEEKS-2006	Weeks
ER40881	HEAD WEEKS LAID OFF-2006	Weeks
ER40882	HEAD UNEMPLOYMENT WEEKS-2006	Weeks
ER40883	HEAD WEEKS OUT OF LABOR FORCE-2006	Weeks
ER40884	WIFE WORK WEEKS-2006	Weeks
ER40885	WIFE WEEKLY WORK HOURS-2006	Hours Worked
ER40886	WF OVERTIME WORK HOURS-2006	Hours Worked

ER40887	WIFE TOTAL HOURS OF WORK-2006	Hours Worked
ER40888	WF WEEKS MISSED FOR ILLNESS OF OTRS-2006	Weeks
ER40889	WF WEEKS MISSED FOR OWN ILLNESS-2006	Weeks
ER40890	WIFE WEEKS OFF FOR VACATION-2006	Weeks
ER40891	WIFE STRIKE WEEKS-2006	Weeks
ER40892	WIFE WEEKS LAID OFF-2006	Weeks
ER40893	WIFE UNEMPLOYMENT WEEKS-2006	Weeks
ER40894	WIFE WEEKS OUT OF LABOR FORCE-2006	Weeks

For weeks and hours variables, there is a three-step process for imputing values:

1. **Pre-imputation Manual Adjustments**
   First, data processing staff applies manual edits to cases where interview notes or other interview information suggests that a value is incorrect. For example, an interviewer might note that a Head/Wife occupation is teacher, that the Head/Wife held only one job and reported 40 weeks of work, with one week of vacation (perhaps the Head/Wife took a trip) and no other time off. The data processing team knows that the Head/Wife didn't report summer time off as vacation time, and so would adjust weeks of vacation.

2. **Imputation**
   After these manual adjustments are performed during the initial data processing steps, Don't Know/NA-Refused values are imputed mechanically.

3. **Post-imputation Manual Adjustments**
   Finally, all hours and weeks values (weeks worked, weeks not worked, hours worked) are reviewed for reasonableness and manually edited if necessary. Once weeks worked and hours worked are calculated, wage rates are calculated using income data (see income section of document) and are edited for reasonableness as well.

9.1.1 Weeks Worked

In the employment section, respondents are asked to report weeks worked on each job a Head or Wife held during 2006. In addition, respondents also report on time spent not working including:

- Weeks not working due to vacation
- Weeks not working due to illness- self
- Weeks not working due to illness – other
- Weeks not working due to strike
- Weeks laid off
- Weeks unemployed

- Weeks out of the labor force

Once pre-imputation manual adjustments are completed by data processing staff, the 2007 interview provides two methods for determining weeks worked for any given job:

1. **Direct:** Respondents are asked to report weeks worked for each job held during 2006 (see ER36168, ER36200, ER36232, ER36264 for Head Weeks, and ER36426, ER36458, ER36490, ER36522 for Wife Weeks). "How many weeks out of the year did (you/Head/Wife) actually work on this job in 2006, not including any time off that you told me about earlier?"

2. **Calculated:** When respondents reply "Every Week Except What Head/Wife Took Off" (97) "Don't Know" (98) or "Refused" (99), the job start and end dates that are reported in the event history calendar (EHC) portion of the interview, together with responses to weeks not working, can be used to calculate weeks worked (See questions BC6/DE6 of the interview). As background, during the EHC portion of the interview, respondents provide detailed information about each Head and Wife job, including the employer name, job title/occupation, industry, job start date, job end date, and time off from work by category.

If we are unable to use the start and end dates to determine weeks worked for a job, we must apply post-imputation manual adjustments and determine the most likely value for weeks worked, given all of the other information the respondent has provided in interview notes. Table 16 below provides a summary of imputations/edits for job-level weeks worked for Heads and Wives. There are six jobs (out of 13,179 Head/Wife jobs) for which we are unable to determine weeks worked using either the weeks worked interview question, or the job start and end date.

Table 16: Weeks Worked Imputation

Number of Head/Wife Jobs	13179
Number of Head/Wife Jobs with Pre-imputation Manual Adjustments Applied	308
Number of Head/Wife Jobs for which we can't calculate Weeks Worked using either the "Direct" or "Indirect" method	6

Once weeks worked are calculated for an individual job, the individual job weeks are used to calculate total weeks worked across all jobs. If a Head/Wife held only

one job during the year, or if the Head/Wife held part year jobs that did not overlap in time, this trivial step involves simply adding up weeks worked across jobs. When a Head/Wife holds only part year jobs that overlap, we use the weeks worked from the job with most weeks worked as the total year weeks worked.

9.1.2 Weeks Not Working

Respondents are also asked to report Head and Wife weeks time-off work for various time-off categories including: own illness, other person illness, vacation, strike and layoff. As with weeks worked, pre-imputation manual adjustments are applied based on information provided in the interview notes and other data provided in the interview. When a respondent indicates that a Head or Wife had time off for one of the above categories, but does not know how much time was taken (98), or refuses to answer (99), the weeks off work are imputed by substituting a constant value. The frequency of both the data processing edits, and the constant value imputation is provided in Table 17.

Table 17: Imputation Values for Time-Off Categories

Variable	Variable Description	Edits Made by Data Processing Staff	Imputations Made Using Constant Substitution	Constant Value Used for Imputation
ER40879	Head Weeks not working due to vacation	148	28	1 Week
ER40878	Head Weeks not working due to illness- self	12	11	0.4 Weeks
ER40877	Head Weeks not working due to illness – other	6	10	0.4 Weeks
ER40880	Head Weeks not working due to strike	0	6	3 Weeks
ER40881	Head Weeks laid off	56	2	2.5 Weeks
ER40890	Wife Weeks not working due to vacation	95	2	1 Week
ER40889	Wife Weeks not working due to illness- self	6	0	0.4 Weeks
ER40888	Wife Weeks not working due to illness – other	5	0	0.4 Weeks
ER40891	Wife Weeks not working due to strike	0	8	3 Weeks
ER40892	Wife Weeks laid off	28	1	2.5 Weeks

9.1.3 Weeks Unemployed and Weeks Out of the Labor Force (OOLF)

In addition to time away from work, respondents are asked to report on both Head/Wife time spent unemployed, but looking for work, and Head/Wife time out of the labor force. Specifically, respondents are asked about unemployment in 2007 PSID family interview questions BC8/DE8 (ER36312/ER36313/ER36316 and ER36570/ER36572/ER36574), "Was there any time last year, in 2006, when Head/Wife was unemployed and looking for work? [IF YES, ASK: When was that?] [IF DK WHEN, ASK: How much time was that in 2006?]. Respondents are asked about time spent out of the labor force in questions BC/DE7 (ER36331/ER36333/ER36335 and ER36589/ER36591/ER36593). "Was there any time in 2005 or 2006 when Head/Wife did not have a job and were not looking for one? [IF YES, ASK: When was that?] [IF DK WHEN, ASK: How much time was that in 2006?]" The first step again in the imputation process is to apply pre-imputation manual adjustments based on interview notes and other information. Of the 1,020 Heads/Wives with non-zero unemployment weeks, 48 require edits, while of the 3610 Heads/Wives with non-zero out of labor force time, 75 require edits.

Next, when respondents report Don't Know (98), or Refused (99) to either weeks unemployed or weeks OOLF, the values are imputed using a set of rules. Essentially, Unemployment and OOLF time is allocated such that weeks worked + time off + weeks unemployed + weeks OOLF sum to 52 weeks. Time-off imputations are summarized in Table 18 below:

Table 18: Time Not Working Imputations

	# Heads/Wives with Weeks Manually Edited	# Heads/Wives with Weeks Adjusted	#Head/Wife with Non-zero Weeks
Unemployment	48	83	1020
OOLF	75	86	3610

9.1.4 Hours Worked

In question BC43/DE43 of the PSID (ER36170/ER36203/ER36234/ER36266 and ER36428/ER36460/ER36492/ER36524), respondents are asked, "On average, how many hours a week did Head/Wife work on this job (this time) in 2006?" As with weeks worked, the first step in imputing average weekly work hours is for the data processing staff to apply pre-imputation manual adjustments. Next, if any respondent reports that Don't Know (998) or NA-Refused (999) for Head/Wife weekly hours, a value of 40 is used for imputation.

Table 19: Work Hours Imputation Summary

Number of Head/Wife Jobs	13179
Number of Head/Wife Jobs for which Hours Worked has had Pre-imputation Manual Adjustments Applied	249
Number of Head/Wife Jobs for which we Impute Using a Value of 40 Hours per Week	172

9.1.5 Overtime Hours Worked

In questions BC45/DE45 of the 2007 PSID (ER36173/er36205/ER36237/ER36270 and ER36431/ER36463/ER36495/ER36527), respondents are asked the following for each Head/Wife job: "How many hours did that overtime amount to in (that period during) 2006?" When a respondent indicates that a Head/Wife worked overtime, but overtime hours are coded as Don't Know (9,998) or NA/Refused (9,999), overtime hours are imputed as 10% of regular hours for the job. Of the 2,252 Head and Wife observations with non-zero overtime hours, we impute overtime hours for 322 Heads/Wives.

Table 20: Overtime Hours Imputation Summary

	Manual Edits	Imputed	Number of Observations with Non-Zero Overtime
Head Overtime (ER46765)	1	243	1665
Wife Overtime (ER46786)	0	78	587

9.2 Wage Rate Calculation

Using corrected (imputed) weekly hours worked, weeks worked and wage and salary income income, an hourly wage rate is calculated.   The hourly wage rate is the calculated by dividing Head Earnings or Wife Earnings (see definition in Table 1 above) by total hours worked.

9.3 Post-Imputation Manual Adjustments

As a final stage in the work weeks, work hours and wage rate processing, variables are checked for consistency and reasonableness, and are hand-edited if necessary. Cases are reviewed for consistency across variables (if wages are positive, weeks and hours must be posistive), as well as for outlier values. Data processing staff reviews flagged cases and edits them as appropriate based on information contained within the interview notes.

10. Summary

Total Family Income is an aggregation of income sources across family members, and to understand income imputation in the PSID, one must review the imputation process of each individual component of Total Family Income. Across all families in the 2007 PSID, the average percent of total family income that is imputed is just under 10% (the median percent is 0%). For those families for whom we impute at least one income source, the average percent of income imputed is 43%. This is not surprising given the fact that Labor Income is the most frequently imputed income source, and would typically account for a large percentage of a family's income. As Table 21 below illustrates, most families have no income sources imputed, while less than 5% of 2007 PSID families have more than two income sources imputed.

Table 21: Number of Families by Number Income Sources Imputed

Number of Income Sources Imputed	Number of 2007 PSID Families	% of Families
0	6398	77.19
1	1154	13.92
2	494	5.96
3	142	1.71
4	63	0.76
5	18	0.22
6	12	0.14
7	4	0.05
8	2	0.02
9	2	0.02

The majority of income sources are imputed using a median substitution method. However, the largest source of income, labor income, is imputed in stages and involves using data from the Employment section of the PSID questionnaire (Sections BC/DE), applying the prior wave's value, and substituting the overall median income amount. Users can identify which observations have been imputed, and by which method, by using the accuracy variables provided in the PSID (see Table 23). The accuracy codes correspond to imputation methods as follows:

Table 22: Income Imputation PSID Codes

Imputation Method	Imputation Code
Imputed from Other Information in the Interview	2
Imputed from Last Wave's Report	3
Imputed from Subgroup Means	4
Imputed Using Median Value of all Non-Zero Cases	5
Hotdeck Replacement	6

Table 23 below provides a summary of income sources, imputation methods used, and the number of cases with valid observations and imputed observations. In addition, the PSID variable that identifies the imputation method is listed in the final column.

Table 23: Income Imputation Summary Table

Source	# Observations with non-zero income (including imputed cases)	# Observations requiring imputations	Methodology	Accuracy Variable
Head Labor Income	6518	544	Step 1, Use PSID Employment Section BC/DE Wages/Hours/Weeks Worked to impute (204), else Step 2, Use prior year income (126), else overall median (214)	ER40904
Wife Labor Income	3101	250	Step 1, Use BC/DE Wages/Hours/Weeks Worked to impute (78), else Step 2, Use prior year income (52), else overall median (120)	ER40934
Head Bonus	585	32	Step 1: Use average bonus percent by OCC code, apply to wages (26), else Step 2, Use overall median bonus amount (6)	ER40906
Head OT	201	60	Step 1, Use BC jobs info (21), else Step 2, avg OT as pct of wages by occ code (18), else median OT amount (21)	ER40908
Head Tips	71	14	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (13), else median tips amount (0)	ER40910
Head Commission	70	14	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (0), else median tips amount (13)	ER40912
Head Professional Practice	92	5	Step 1, Use Prior year (1), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (4)	ER40914
Head Market Gardening	34	3	Step 1, Use Prior year (0), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (3)	ER40916
Head Roomers and Boarders	0	0	Step 1, Use Prior year (0), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (0)	N/A

Head Farm Income	35	6	Step 1, Farm Receipts - Farm Expenses (0), else Step 2, Prior Year Income (5) else Step 3, overall median farm income (1)	ER40899
Head Business	337	85	Step 1, Use BC/DE Self Employment Income (10), else Step 2, Prior year's income if same industry (2), else Step 3, hot deck within industry (73)	ER40896
Wife Business	110	22	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (22)	ER40896
Head & Wife Business	74	25	Step 1, Use BC/DE Self Employment Income (3), else Step 2, Prior year's income if same industry (3), else Step 3, hot deck within industry (19)	ER40896
OFUM Only Business	6	7	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (7)	ER40896
Head & OFUM Business	2	0	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (0)	ER40896
Head, N/A Business	7	6	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (6)	ER40896
Head, Wife, N/A Business	2	0	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (0)	ER40896
Head, OFUM, N/A Business	0	1	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (1)	ER40896
N/A Business	1	0	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (0)	ER40896
Rent Head	486	13	Overall Median	ER40923
Dividend	1211	182	Overall Median	ER40925

Head				
Interest Head	2374	326	Overall Median	ER40927
Trust Head	81	8	Overall Median	ER40929
Rent Wife	182	6	Overall Median	ER40936
Dividends Wife	641	95	Overall Median	ER40938
Interest Wife	1325	165	Overall Median	ER40940
Trust Wife	13	0	Overall Median	ER40942
OFUM Job 1	2352	756	Overall Median (within Job #)	ER40996
OFUM Job 2	452	156	Overall Median (within Job #)	ER40996
OFUM Job 3	98	37	Overall Median (within Job #)	ER40996
OFUM Job 4	29	15	Overall Median (within Job #)	ER40996
OFUM Interest	80	8	Overall Median	ER40998
OFUM ADC	19	4	Overall Median	ER41001
OFUM SSI	191	9	Overall Median	ER41003
OFUM Welfare	28	3	Overall Median	ER41005
OFUM Veterans	8	4	Overall Median	ER41007
OFUM Pension	61	20	Overall Median	ER41009
OFUM Unemployment	10	0	Overall Median	ER41011
OFUM Workers Comp	4	0	Overall Median	ER41013
OFUM Child Support	28	7	Overall Median	ER41015
OFUM Relatives	35	3	Overall Median	ER41017
OFUM Other	145	20	Overall Median	ER41019
Social Security	2244	176	Overall Median	ER41022, ER41024, ER41026
Head Alimony	38	0	Overall Median	ER40965
Head Annuity	147	15	Overall Median	ER40955
Head Child Support	417	13	Overall Median	ER40963

Head Help Non Rel	443	40	Overall Median	ER40969
Head Help Rel	962	76	Overall Median	ER40967
Head Other	221	18	Overall Median	ER40971
Head Retirement	772	39	Overall Median	ER40957
Head SSI	218	11	Overall Median	ER40947
Head TANF	129	2	Overall Median	ER40945
Head Unemp	333	13	Overall Median	ER40959
Head VA Pension	211	16	Overall Median	ER40951
Head Welfare	46	4	Overall Median	ER40949
Head Workers Comp	84	5	Overall Median	ER40961
Wife ADC	10	1	Overall Median	ER40973
Wife Child Support	146	7	Overall Median	ER40985
Wife Help Non Rel	12	2	Overall Median	ER40989
Wife Help Rel	151	13	Overall Median	ER40987
Wife Other	138	8	Overall Median	ER40991
Wife Pension	206	8	Overall Median	ER40979
Wife SSI	43	1	Overall Median	ER40975
Wife Unemp	118	5	Overall Median	ER40981
Wife Welfare	8	1	Overall Median	ER40977
Wife Workers Comp	22	1	Overall Median	ER40983