HRS DATA ON INTERGENERATIONAL TRANSFERS

REVISION FEB 2008

last update: February 27, 2008

DATA PREPARATION DESCRIPTION

Note: Reported numbers of observation and variables in intermediary data files are indications only.

---

1 Overview of Procedure

1.1 Introduction

We use data from the Health and Retirement Study (HRS), available online via the University of Michigan's Institute for Social Research web site. In addition, we use data from the RANDHRS variable collection, which contains derived variables from the original HRS distribution. The RANDHRS distribution is made available from RAND, also via the HRS web site. The HRS website has last been accessed for checking data alerts and possible download of data on November 11, 2005. For replication of this work, we cannot assume responsibility for changes to data or file structure made after that date.

The data is a panel data set with biennual sampling and contains information on older Americans. The survey was started in 1992, the latest published data available are from the 2002 wave when revision of the data started (preliminary data for 2004 were also available, but not used). There are several birth cohorts in the sampled population, and we focus here on the HRS Subsample, households where at least one respondent was born in between the years 1931 and 1941. These people are followed over time in subsequent interviews. There is no refreshment sampling for respondents that drop out. Respondents leave the sample due to death or due to refusal to participate.

The main files are in the 'core' of the distribution, containing all regular interviews. In addition, exit interviews are being held in households where one of the main respondent has passed away between waves. The exit interview is typically held with the surviving spouse or some other 'proxy' interviewee. This data review disregards exit interviews. Some crucial variables have been imputed by HRS staff and (single) imputations are distributed in imputation files. Of interest as well are imputations on incomes, wealth, and transfers to children are of interest, although we also provide own imputations, or rely on RANDHRS files. To link data across time for the same respondents, 'tracker' files are provided by the HRS project team. These tracker files come in two versions: one respondent level file, and one 'other person' level distribution, LOPN.

The HRS data staff makes files available in 'versions'. These contain changes to the original data in the sense of cleaning and consistency checking done. Occasionally, improvements are published on the web site, and new versions are being made available. In addition to such data updates, frequently the data provider issues 'data alerts' that report on errors and omissions in the data and appropriate fixes. We use the most recent data versions available and update where necessary taking into account data alerts.

The data are suitable to estimate economic models based on observed behavior of intrafamily monetary transfers. For this, we need information on a parental household and their children. The information needed consists of

1. identifiers for all households, parents and children in the household, consistent over time
2. individual characteristics of parents and children

3. indicators of whether or not transfers were made between parents and children in a given wave

4. amounts of transfers made in a given year

While conceptually straightforward, collating this information is associated with a lot of work due to remaining inconsistencies, lack of appropriate documentation, and problems with merging the various pieces of data. This document serves to help keeping track of checks and changes that have been made and what procedures have been applied.

1.2 Identifiers

HRS data, data sets, but also some identifiers and other variables are sometimes following a naming convention which is informative on the particular wave they relate to. There are six waves which are labeled with a letter: `A`, `C`, `E`, `F`, `G`, and `H`, referring to waves 1992-2002. We sometimes refer to those waves as waves 1 through 6.

There are four (or five) levels of identifiers in the data, information which is taken from the two tracker files (respondent and other-person tracker files):

1. A household identifier, `HHID`

2. A wave-specific sub-household identifier, say `CSUBHH` in wave 2 ('`C`'), which becomes relevant when an existing household splits in two, say. A household unit is therefore identified by `HHID xSUBHH`. `xSUBHH` takes typically a value between 0 and 2, but assumes the values 3 or 4 when a household member has died between waves, and value 9 when a household member is not yet present (but joins in a later wave, such as a new spouse), or when it has died a while ago; values 3 and 4 typically also indicate (candidates for) exit interviews. In the first wave, `ASUBHH`, is zero for all (except for spouses that join later, '9').

3. A person-level identifier, typically a respondent's person number, `PN`, within the household, or, a 'longitudinal other person number', `LOPN`. `LOPN` identifies non-respondents (such as children) over time (survey waves). Since the information on a `LOPN` case is provided by a main respondent, and since households may split, this may generate more than one respondent reporting on the same `LOPN` (after the first wave).

4. In order to identify the source of the information, HRS staff introduced a link identifier, `STORY`. It *tends* to associate a particular `LOPN` with a particular `PN`. There are typically one, but up to three different 'stories' in the data.[1] Note: `LOPN` are originally only available in a single file that collects identifying information on all non-respondents. However, especially in wave one (and occasionally in other waves), respondents report on children that do not receive a `LOPN` identifier. We exclude all such observations from further analysis. We devise in addition a new identifier that we refer to as `FAMSTORY`. It serves a similar purpose as `STORY` but conditions on the identity of the reporting family respondent rather than on the identity of the reporting (sub-)household, within which there may be a change of family respondent over time. This is explained in more detail below.

5. A sub-person-level identifier, typically a numbered item on a list; this applies for instance to transfers made from parents to children when there are several transfers made to the same child.

Data sets listed below give the identifier which allow unique identification and merging of data sets across waves. Not all original data sets carry all identifiers. Part of the work was in figuring out how to associate various observations with what identifiers.

[1] Suppose there are initially two respondents in a household, and the household splits in two after the first wave, then both sub-households may have their own family respondent who delivers a 'story' on the child (or other person) in question.

1.3 Monetary Values

All monetary values relating to wealth, income, and transfers, have been converted into 1991 dollars using the 'All Urban Consumers' Consumer Price Index as published on the Bureau of Labor Statistics' website (series ID : CUUR0000SA0), rebased where necessary. The conversion factors are

CPI1991	1.0000	HRS 1992
CPI1993	1.0609	HRS 1994
CPI1995	1.1189	HRS 1996
CPI1997	1.1784	HRS 1998
CPI1999	1.2232	HRS 2000
CPI2001	1.3003	HRS 2002
CPI2003	1.3510	HRS 2004

source: BLS website, Nov 2005

1.4 Initial changes to data: `inidata.do`

This file applies certain changes to the original data, following documentation, in particular the LOPN project (see: `LOPNDD.pdf`, distributed by HRS). The LOPN project description that we rely on, stems from February 2004. After that date a number of changes have been made to the original data, rendering some of the suggested changes obsolete. There are a number of LOPN files specific for each wave containing `LOPN` identifiers. The file `inidata.do` attaches `LOPN` to a number of original files, and creates new children level files, (names with 'rev'—revised version—appended) according to the following table:

Table 1: File Renaming after LOPN

wave	DATA SET	DISTR	LOPN DATA	identifier	Nobs
1992	HHLIST_REV	HRS core	LC92_HL	HHID ASUBHH OPN	7,194
	KIDS_REV	HRS core	LC92_K	HHID ASUBHH OPN	24,452
1994	W2HHLIST_REV	HRS core	LCX94_HL	HHID CSUBHH OPN	8,417
	W2KIDS_REV	HRS core & exit	LCX94_K	HHID CSUBHH OPN	22,861
1996	H96PR_MC_REV	HRS core	LC96_MC	HHID ESUBHH OPN	25,171
	H96D_MC_REV	HRS core	LC96_MC	HHID ESUBHH OPN	25,171
	X96PR_MC_REV	HRS exit	LX96_MC	HHID PSUBHH OPN	976
1998	H98PR_MC_REV	HRS core	LC98_MC	HHID FSUBHH OPN	25,152
	H98D_MC_REV	HRS core	LC98_MC	HHID FSUBHH OPN	25,152
	X98PR_MC_REV	HRS exit	LX98_MC	HHID FSUBHH OPN	1,075
2000	H00PR_MC_REV	HRS core	LC00_MC_orig	HHID GSUBHH OPN	24,195
	H00D_MC_REV	HRS core	LC00_MC_orig	HHID GSUBHH OPN	24,195
	X00PR_MC_REV	HRS exit	LX00_MC	HHID RSUBHH OPN	1,266
2002	H02PR_MC_REV	HRS core	LC02_MC_orig	HHID HSUBHH OPN	37,921
	H02E_MC_REV	HRS core	LC02_MC_orig	HHID HSUBHH OPN	37,921
	X02PR_MC_REV	HRS exit	LX02_MC	HHID SSUBHH OPN	2,745

* = new name, after revision according to LOPNDD.PDF and merging in LOPN

In addition, the file changes in some (especially imputation data) data sets numerical into a string identifiers.

1.5 Main changes to data: `dataprep.do`

The main file for data preparation (`dataprep.do`) goes through a number of steps (invoked as separate files, `dataprep*.inc`), all producing intermediary data files which are used as inputs at some later stage. We list the main sections of this file along with a short description of the contents and a listing of all relevant input and output files.

SECTION I: PARENT-LEVEL IDENTIFIERS FROM HRS TRACKER

File: `dataprep_parentid.inc`.

Collects basic fixed characteristics of respondent and spouse, including their relationship (i.e., which person to match with whom at what wave). Uses this information on respondent level to assess whether or not household composition has changed over time. This gives rise to three possible definitions on whether a household is 'intact'.

According to the first definition, a household is 'intact' either if it contains only one respondent over time, or if it contains two respondents who are and stay a couple. This definition does not condition on whether a respondent dies, unless the household is a couple where only one of the two members dies between the waves. Not intact, according to this definition are couples that split into sub-households, or where a new partner enters.

A second definition conditions on survival between waves, but is otherwise the same as the first definition. A third definition simply indicates a change of sub-household identifier between the waves. It may seem reasonable to apply the second definition.

Note that 'intact' here does not mean that the sub-household has given an interview in the respective year. That implies that even though households are counted as intact, their information may be missing in a particular wave. The following table gives response rates (relative frequencies of interviews conducted) at sub-household level.

HRS 1992	100.00
HRS 1994	91.46
HRS 1996	88.61
HRS 1998	88.06
HRS 2000	86.26
HRS 2002	87.60

Finally, the file determines person-level identifier of family and financial respondents.

Table 2: SECTION I FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
trk2002	input	27,107	213	HHID PN	1992-2002	orig.	respondent-level identifiers and fixed demographics, background characterstics, and interview information; covers all HRS-related samples, including AHEAD, etc.
NDI2000	input	4,114	11	HHID PN	1992-2000	orig.	(National Death Index) respondent-level indicator on death and death probabilities; covers all HRS-related samples, including AHEAD, etc.
Region	input	26,935	30	HHID PN	1992-2002	orig.	records region of residence and degree of urbanization per respondent and wave, also indicates miles moved between two survey waves; covers all HRS-related samples
kids_rev	input	24,452	49	HHID LOPN	1992	orig.	children-level information
w2kids_rev	input	22,861	44	HHID LOPN	1994	orig.	children-level information
h96pr_mc_rev	input	25,171	26	HHID LOPN	1996	orig.	children-level information

Table 2: (cont'd.)

file name	level	Nobs	Nvars	identifier	wave	status	remarks
h98pr_mc_rev	input	25,152	36	HHID LOPN	1998	orig.	children-level information
h00pr_mc_rev	input	24,195	34	HHID LOPN	2000	orig.	children-level information
h02pr_mc_rev	input	37,921	28	HHID LOPN	2002	orig.	children-level information
kidsI1	output	24,452	50	HHID LOPN	1992	interm.	children-level information
kidsI2	output	22,861	45	HHID LOPN	1994	interm.	children-level information
kidsI3	output	25,171	27	HHID LOPN	1996	interm.	children-level information
kidsI4	output	25,152	37	HHID LOPN	1998	interm.	children-level information
kidsI5	output	24,195	35	HHID LOPN	2000	interm.	children-level information
kidsI6	output	37,921	29	HHID LOPN	2002	interm.	children-level information
tracktmpI	output	13,265	255	HHID PN	1992-2002	interm.	basically tracker file, records whether a household is 'intact' at any of the three definitions described

SECTION II: CHILDREN-LEVEL STORY INDICATORS

File: `dataprep_story.inc`.

Uses LOPN-tracker file (longitudinal other person number) and extracts the `STORY` identifier per 'child'. At this level, the file contains not only children but a host of other non-respondent persons.

As it turns out, `STORY` obtains from sorting on `HHID` and all wave-specific `SUBHH` ID's (including those for exit interviews). With further assumptions (the procedural definition is undocumented in the HRS distribution), one can match `STORY` with other files. `STORY` is extracted and pasted on wave-specific LOPN-level files.

The `STORY` indicator should be useful in merging data sets to tell which information has been supplied by which respondent over time. Note, though, that there are cases where, for a given `STORY`, the family respondent does change over time. Where it happens, though, such a change is not associated with a change in sub-household. That means, if in an intact household in one wave the father acts as family respondent and in the next wave the mother, the `STORY` indicator would not change since it relates to the same household. Therefore, there is remaining variation in terms of children characteristics within a given 'story'. In order to control for this, we introduced an additional linker, `FAMSTORY` that keeps the `PN`/`LOPN` association constant over time. In essence, `FAMSTORY` identifies within `HHID LOPN` the source of the information given.

The children-level files will now contain slightly fewer numbers of observations since we drop those where no match could be established with a `LOPN`.

Table 3: SECTION II FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
HRS9202	input	54,598	26	HHID LOPN STORY	1992-2002	orig.	'other person'-'story'-level indicator to determine which 'story' was told in which wave in which subhousehold
kidsI1-kidsI6	input						see Section I
kidsII1	output	24,452	63	HHID LOPN FAMSTORY	1992	interm.	children-level information
kidsII2	output	22,856	57	HHID LOPN FAMSTORY	1994	interm.	children-level information
kidsII3	output	25,158	39	HHID LOPN FAMSTORY	1996	interm.	children-level information
kidsII4	output	25,128	49	HHID LOPN FAMSTORY	1998	interm.	children-level information
kidsII5	output	24,182	47	HHID LOPN FAMSTORY	2000	interm.	children-level information
kidsII6	output	37,896	41	HHID LOPN FAMSTORY	2002	interm.	children-level information

SECTION III: ALIGN IDENTIFIERS CHILDREN/PARENT LEVEL

File: `dataprep_parentkidid.inc`.

Combines respondent-level and children level files in order to get various respondent-level `PN` identifiers (for family and financial respondents) into accordance. Reason: there is occasionally disagreement, or `PN` identifiers are missing in one of the file levels.

In addition, determine the `PN` of the respondent who will be interpreted as either family or financial respondent of the 'household', even though this household may split up in later waves or change family respondent in the course of time. Fixing the identity of a family or financial respondent helps avoiding unwanted changes in characteristics over time. The family respondent or financial respondent are typically those of the first wave, and will be substituted with some other respondent if not available in the first wave. Note: not all households have a financial or family respondent (even though children may be present). There are typically fewer financial respondents available than family respondents. The number of both financial and family respondents in the sample decreases over time.

Table 4: SECTION III FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
tracktmpI	input						see Section I
kidsII1-kidsII6	input						see Section II
tracktmpIII	output	13,265	245	HHID PN	1992-2002	interm.	respondent-level information
kidsIII1	output	24,452	66	HHID LOPN FAMSTORY	1992	interm.	children-level information
kidsIII2	output	22,856	60	HHID LOPN FAMSTORY	1994	interm.	children-level information
kidsIII3	output	25,158	42	HHID LOPN FAMSTORY	1996	interm.	children-level information
kidsIII4	output	25,128	52	HHID LOPN FAMSTORY	1998	interm.	children-level information
kidsIII5	output	24,182	50	HHID LOPN FAMSTORY	2000	interm.	children-level information
kidsIII6	output	37,896	44	HHID LOPN FAMSTORY	2002	interm.	children-level information

SECTION IV: CLEAN UP CHILD STATUS BETWEEN WAVES

File: `dataprep_definekid.inc`.

Determine the relationship between a respondent (or the spouse) and the 'other person'. In most waves, the 'other person' will be a child of at least one of the main respondents, but sometimes the person in question is either not a child to (any of) the (two) main respondent(s) or a child to only one of them (and a step-child to the other).

The information has been elicited using different types of questions over the years, leading to structural breaks especially between waves 1 and 2 and 2 and 3. Our file cleans up inconsistencies and fills in gaps. Typically, as interviewees are being presented with previous interview data, the more recent information may be more accurate and serve as an update. That is the assumption made here.

We supply two indicators, telling whether a person is or is not a child to the household, and whether it is a current respondent's or current spouse's 'own child' or 'step child'. Due to change of family respondent or financial respondent, this 'relationship indicator' may seem to change over time within a given household or subhousehold. Use of the link variable `FAMSTORY` eliminates such variation.

Remark: the internal use of the label 'natural' actually refers to respondent's own child as opposed to step-child and not to adoptees. (There are a total of 110 adoptees in the first wave mentioned, corresponding to 0.45% of all children mentioned).

Table 5: SECTION IV FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
kidsIII1-kidsIII6	input						see Section III
kidsIV1	output	24,452	77	HHID LOPN FAMSTORY	1992	interm.	children-level information
kidsIV2	output	22,856	71	HHID LOPN FAMSTORY	1994	interm.	children-level information
kidsIV3	output	25,158	53	HHID LOPN FAMSTORY	1996	interm.	children-level information
kidsIV4	output	25,128	63	HHID LOPN FAMSTORY	1998	interm.	children-level information
kidsIV5	output	24,182	61	HHID LOPN FAMSTORY	2000	interm.	children-level information
kidsIV6	output	37,896	55	HHID LOPN FAMSTORY	2002	interm.	children-level information

SECTION V: CHILD-LEVEL INFORMATION ON BEING INCLUDED OVER TIME

File: `dataprep_includekid.inc`.

Excludes those LOPN from consideration that are not a child of either family respondent or the spouse (i.e., not a grandchild, in-law, etc.). This explains the drop in observations. Combines all child-level identifiers into a single file, k_id1. This file only contains identifiers, but no substantial information useful for further analysis.

Table 6: SECTION V FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
tracktmpIII	input						see Section III
kidsIV1-kidsIV6	input						see Section IV
kidsV1	output	24,400	77	HHID LOPN FAMSTORY	1992	interm.	children-level information
kidsV2	output	22,830	71	HHID LOPN FAMSTORY	1994	interm.	children-level information
kidsV3	output	22,856	53	HHID LOPN FAMSTORY	1996	interm.	children-level information
kidsV4	output	22,883	63	HHID LOPN FAMSTORY	1998	interm.	children-level information
kidsV5	output	22,051	61	HHID LOPN FAMSTORY	2000	interm.	children-level information
kidsV6	output	22,001	55	HHID LOPN FAMSTORY	2002	interm.	children-level information
k_id1	output	32,081	116	HHID LOPN FAMSTORY	1992-2002	interm.	children-level identifiers

SECTION VI: DEFINITION OF CHILDREN-INTACT SUB-HOUSEHOLDS

File: `dataprep_kidintact.inc`.

Suppose, we condition on a sub-household, then over time the number of children mentioned can change by either children not mentioned before 'entering' or children mentioned before not being mentioned anymore. There are very few occasions where a child is born into a household or dies between waves. It is more likely that a new spouse entering the household also has children of his or her own.

This requires to define whether a sub-household is intact at the child level and information on parent/subhousehold level. We employ two definitions of being intact between waves at the child level: the first disallows any splitting of the household into sub-households, the second conditions on sub-households and counts those whose composition is unchanged at child level.

The resulting file, k_id2, does not contain substantive variables.

Table 7: SECTION VI FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
k_id1	input						see Section V
k_id2	output	32,081	135	HHID LOPN FAMSTORY	1992-2002	interm.	children-level identifiers; information on whether a (sub-)household is 'intact' at the child level

SECTION VII: PARENTAL CHARACTERISTICS

File: `dataprep_parentchar.inc`.

Extracts respondent-level characteristics from three types of sources. Characteristics can refer to a particular respondent or to the entire (sub-)household. Fixed characteristics, such as age, sex, education etc., plus some other background characteristics, can originate from the Tracker file. Health, income and wealth information is from wave-specific files.

Note that both income and wealth are composites of a variety of sources, on which missing values tend to aggregate. Therefore, it is advisable to use imputed aggregates. HRS staff does provide both income and wealth imputations. It turns out, however, that there are structural breaks over time in the series (for instance, wealth in wave one is on average about twice the average wealth in wave 2). Such and other inconsistencies are avoided by using imputed income and wealth aggregates from the RANDHRS files. The difference between the HRS and the RANDHRS imputations is that the former is simple hotdeck (conditional on bracket information), while the latter also takes into account (cross-sectional) variates and predicts from a regression.

Output files come in three versions: variables per wave, 'next to each other'; panel data on respondent level; panel data on sub-households level, distinguishing family respondent characteristics from those of the financial respondent.

Table 8: SECTION VII FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
tracktmpIII	input						see Section III
health	input	12,652	791	HHID PN	1992	orig.	respondent-level health
househld	input	7,607	1,032	HHID ASUBHH	1992	orig.	household-level wealth and income
w2b	input	11,596	183	HHID PN	1994	orig.	respondent-level health
h94i_n_rev	input	6,979	301	HHID CSUBHH	1994	orig.	household-level income
h94i_k_rev	input	6,979	82	HHID CSUBHH	1994	orig.	household-level wealth
h96b_r	input	10,964	186	HHID PN	1996	orig.	respondent-level health
h96i_jh	input	6,816	484	HHID ESUBHH	1996	orig.	household-level wealth and income
h98b_r	input	21,384	190	HHID PN	1998	orig.	respondent-level health
h98i_jh	input	14,395	416	HHID FSUBHH	1998	orig.	household-level wealth and income
h00b_r	input	19,579	190	HHID PN	2000	orig.	respondent-level health
h00i_jh	input	13,214	416	HHID GSUBHH	2000	orig.	household-level wealth and income
h02c_r	input	18,166	195	HHID PN	2002	orig.	respondent-level health
h02i_qh	input	12,350	424	HHID HSUBHH	2002	orig.	household-level wealth and income
rndhrs1e7	input	26,728	794	hhidpn	1992	orig.	household-level wealth and income
rndhrs2e7	input	26,728	903	hhidpn	1994	orig.	household-level wealth and income
rndhrs3e7	input	26,728	866	hhidpn	1996	orig.	household-level wealth and income
rndhrs4e7	input	26,728	866	hhidpn	1998	orig.	household-level wealth and income
rndhrs5e7	input	26,728	868	hhidpn	2000	orig.	household-level wealth and income
rndhrs6e7	input	26,728	879	hhidpn	2002	orig.	household-level wealth and income
parnt	output	13,265	142	HHID PN	1992-2002	interm.	respondent-level characteristics, wave-specific variables
parnts	output	65,250	52	HHID PN wave	1992-2002	interm.	respondent-level characteristics, panel data

Table 8: (cont'd.)

file name	level	Nobs	Nvars	identifier	wave	status	remarks
p12tmp	output	40,658	89	HHID wave	1992-2002	interm.	characteristics for family and financial respondents, sub-household level

SECTION VIII: TRANSFERS TO CHILDREN: PREPARE INDICATORS

File: `dataprep_preptc.inc`.

Preparatory section to determine transfers to children. From wave 3 on, transfers could be mentioned in the form of 'same amount to all my children/grandchildren'. This section extracts, among others, the child-level identifiers for those cases to help associate transfer amounts to individual children. Note that one respondent could indicate several transfers to the same child, including 'same amount' transfers to all children in addition to individual transfers. Therefore, the set of all children needs to be identified for imputation.

In addition, the file extracts some demographic information at the child level for waves 3 and above.

Table 9: SECTION VIII FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
k_id2	input						see Section VI
h96d_h	input	6,816	492	HHID ESUBHH	1996	orig.	household-level any transfers
h96d_tc	input	3,901	22	HHID ESUBHH OPN ETYPTRAN	1996	orig.	transfer-level
h98d_h	input	14,395	523	HHID FSUBHH	1998	orig.	household-level any transfers
h98d_tc	input	6,176	27	HHID FSUBHH OPN FTRANNUM	1998	orig.	transfer-level
h00d_h	input	13,214	545	HHID GSUBHH	2000	orig.	household-level any transfers
h00d_tc	input	5,839	27	HHID GSUBHH OPN GTRANNUM	2000	orig.	transfer-level
h02e_h	input	12,350	108	HHID HSUBHH	2002	orig.	household-level any transfers
h02e_tc	input	5,133	24	HHID HSUBHH OPN HTC_NDX	2002	orig.	transfer-level
tmp3_38	output	757	4	HHID LOPN	1996	interm.	'all children' identifiers
tmp4_38	output	673	4	HHID LOPN	1998	interm.	'all children' identifiers
tmp5_38	output	684	4	HHID LOPN	2000	interm.	'all children' identifiers
tmp6_38	output	545	4	HHID LOPN	2002	interm.	'all children' identifiers
tc96	output	3,832	26	HHID ESUBHH OPN ETYPTRAN	1996	interm.	transfers to children
tc98	output	3,128	31	HHID FSUBHH OPN FTRANNUM	1998	interm.	transfers to children
tc00	output	2,884	31	HHID GSUBHH OPN GTRANNUM	2000	interm.	transfers to children
tc02	output	2,507	28	HHID HSUBHH OPN HTC_NDX	2002	interm.	transfers to children
tmpkds3	output	22,856	5	HHID ESUBHH OPN2	1996	interm.	children ID's
tmpkds4	output	22,883	5	HHID FSUBHH OPN2	1998	interm.	children ID's
tmpkds5	output	22,051	5	HHID GSUBHH OPN2	2000	interm.	children ID's
tmpkds6	output	22,001	5	HHID HSUBHH OPN2	2002	interm.	children ID's
tmpdemo4	output	23,666	9	HHID FSUBHH OPN	1998	interm.	children demographics

Table 9: (cont'd.)

file name	level	Nobs	Nvars	identifier		wave	status	remarks
tmpdemo5	output	22,548	9	HHID OPN	GSUBHH	2000	interm.	children demographics
tmpdemo6	output	22,002	9	HHID OPN	HSUBHH	2002	interm.	children demographics

SECTION IX: EXTRACT DEMOGRAPHICS AND TRANSFER AMOUNTS, IMPUTE TRANSFERS

File: `dataprep_kiddemotc.inc`.

Extracts demographics and other children characteristics from wave-specific files. Extracts information on whether transfers were made to the child and how much. It also uses the information on transfers to all children and assigns them to each child. Many amounts are only given in brackets (if at all), and we may want to use imputed values. Imputations are available from the HRS (imputed data sections). Since the imputed amounts are of crucial importance for the current project, we impute the values ourselves (hotdeck conditional on bracket information) to iron out inconsistencies in the HRS imputations and to have multiple imputations available.

Table 10: SECTION IX FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
kidsV1-kidsV6	input						see Section V
tmp3_38-tmp6_38	input						see Section VIII
tc96-tc02	input						see Section VIII
k_id2	input						see Section VI
tmpkds3..6	input						see Section VIII
tmpdemo4..6	input						see Section VIII
health	input	12,652	791	HHID ASUBHH	1992	orig.	parent-level data, child-level info
h94i_kid_rev	input	22,861	24	HHID CSUBHH OPN	1994	orig.	children-level information
h96d_mc_rev	input	25,171	41	HHID ESUBHH OPN	1996	orig.	children-level information
h96i_dtc	input	3,901	11	HHID ESUBHH OPN ETYPTRAN	1996	orig.	transfer-level
h98d_mc_rev	input	25,152	49	HHID FSUBHH OPN	1998	orig.	children-level information
h98i_dtc	input	6,176	11	HHID FSUBHH OPN FTRANNUM	1998	orig.	transfer-level
h00d_mc_rev	input	24,195	49	HHID GSUBHH OPN	2000	orig.	children-level information
h00i_dtc	input	5,839	11	HHID GSUBHH OPN GTRANNUM	2000	orig.	transfer-level
h02e_mc_rev	input	35,413	35	HHID HSUBHH OPN	2002	orig.	children-level information
h02i_etc	input	5,133	9	HHID HSUBHH OPN HTC_NDX	2002	orig.	transfer-level
Kamttmp	output	137,021	19	HHID LOPN FAMSTORY wave	1992-2002	interm.	children-level transfers
Kchartmp	output	32,081	222	HHID LOPN FAMSTORY	1992-2002	interm.	children-level characteristics

SECTION X: COMPARE CHILDREN-LEVEL CHARACTERISTICS ACROSS WAVES

File: `dataprep_compkid.inc`.

This section cleans demographics and background characteristics for children across waves. Variation in children characteristics can come from many sources. One source is that certain variables were not elicited in all waves, or only from part of the respondents. In this case, missing values need to be filled in. In addition, there are recall errors or other types of response errors that lead a certain respondent to characterize their children differently over time even in terms of fixed characeristics (such as sex and year of birth). We correct 'errors' as far as possible, yielding sex and year of birth information that does not change over time for a given child. Education is made to be nondecreasing over time, and having children in two waves but not in a wave in between is also interpreted as error (we do allow for children of a child to die or to be born, though). This assumption seems sensible as child mortality is typically low in the US.

Other characteristics, such as whether or not a child attends school, stays at home, works, or its income, are not updated.

Table 11: SECTION X FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
Kchartmp	input						see Section IX
Ktmp	output	136,797	121	HHID LOPN FAMSTORY wave	1992-2002	interm.	children-level identifiers

SECTION XI: CHILDREN: PREPARE DATA COMBINATION

File: `dataprep_prepcomb.inc`.

This section is an auxiliary section to prepare for section XII. Among others, it determines the `PN` ID of both father and mother of the children. From here, we determine the identity of the family and financial respondent.

SECTION XII: COMBINE DATA

File: `dataprep_combine.inc`.

Finally, combine all data: identifiers, parental characteristics for family and financial respondent, children's characteristics, and transfers made from parents to children, resulting in a panel data set tracking all children over the entire six waves and their associated parental household characteristics, including all 'stories' told about those children.

Also define bracket values per wave for transfer amounts and deflate.

Table 12: SECTION XII FILES

file name	level	Nobs	Nvars	identifier	wave	status	remarks
Ktmp_pinc	input					interm.	see Section XI
p12tmp	input					interm.	see Section VII
Kamttmp	input					interm.	see Section IX
all	output	136,797	243	HHID LOPN wave FAMSTORY	1992-2002	final	all relevant information, panel data set

2 List of Files

2.1 Procedure and Stata Files

To recreate the dataset used for analysis, the files `inidata.do` and `dataprep.do` (including all input files, `.inc`) need to be run.

Table 13: Stata Files

run order	file name	content
		Stata .do files
1	dataalert.do	change data, taking into account alerts
2	inidata.do	prepare data, see LOPNDD.pdf
3	dataprep.do	data preparation (frame)
i	parentid.inc	parent-level indicators
ii	story.inc	story-level indicator
iii	parentkidid.inc	align parent/kid indic.
iv	definekid.inc	define child status
v	includekid.inc	include children
vi	kidintact.inc	hh intact at child level
vii	parentchar.inc	parental characteristics
viii	preptc.inc	prepare transfers
ix	kiddemotc.inc	child demographics & transfers
x	compkid.inc	compare child demographics
xi	prepcomb.inc	prepare data combination
xii	combine.inc	combine various files into final output file, `all.dta`
		documentation files
	dataalert.txt	list of data alerts
	datades.tex	this file

2.2 HRS Data Distribution

Table 14: Status of Current Data Sets (last update: 09-Dec-2005)

wave	version	type	release date
HRS1992			
.. CORE	2.0	FINAL	200409
.. IMP	3.0	FINAL	200311
.. EXIMP			
HRS1994			
.. CORE & EXIT	2.0	FINAL	200409
.. IMP	3.0	FINAL	200311
HRS1996			
.. CORE	4.0	FINAL	200309
.. EXIT	1.0	FINAL	200309
.. IMP	3.0	FINAL	200212
.. EXIMP	1.0	FINAL	200406
HRS1998			
.. CORE	2.3	FINAL	200311
.. EXIT	1.0	FINAL	200501
.. IMP	3.0	FINAL	200212
.. EXIMP	2.0	EARLY	200406

Table 14: (cont'd.)

wave	version	type	release date
HRS2000			
.. CORE	1.0	FINAL	200209
.. EXIT	1.0	FINAL	200503
.. IMP	2.0	FINAL	200212
.. EXIMP	1.0	EARLY	200209
HRS2002			
.. CORE	1.0	FINAL	20040924
.. EXIT	1.0	EARLY	200402
.. IMP	1.0	FINAL	20051025
TRACKER			
.. REGION	1.0	—	20040203
.. NDI2000	1.0	—	20030224
.. TRK2002	2.0	—	20051026
LOPN			
.. HRS9202	1.1	—	20040217
RANDHRS			
.. RNDHRS	E	—	200508

2.3 List of Original Data Files

Table 15: List of Original Data Files

filename	distribution	wave	ID	Nobs	Nvar	content & remarks
trk2002	HRS TRACKER	1992-2002	HHID PN	27,107	213	longitudinal resp. & household; covers all HRS-related samples (inc. AHEAD etc.)
NDI2000	HRS NDI	1992-2000	HHID PN	4,114	11	National Death Index, respondents; indicator: death and death probabilities
Region	HRS REGION	1992-2002	HHID PN	26,935	30	region of residence, respondents; inc. urbanization, distance of relocation
hrs9202	HRS LOPN	1992-2002	HHID LOPN STORY	54,598	26	LOPN ID
LC92_HL	HRS LOPN	1992	HHID ASUBHH OPN	7,237	4	LOPN ID; household listing
LC92_K	HRS LOPN	1992	HHID ASUBHH OPN	24,697	4	LOPN ID; children
LCX94_HL	HRS LOPN	1994	HHID CSUBHH OPN	8,417	4	LOPN ID; household listing (inc. exit)
LCX94_K	HRS LOPN	1994	HHID CSUBHH OPN	22,861	4	LOPN ID; children (inc. exit)
LC96_MC	HRS LOPN	1996	HHID ESUBHH OPN	25,171	4	LOPN ID; children
LX96_MC	HRS LOPN	1996	HHID PSUBHH OPN	976	4	LOPN ID; children (exit)
LC98_MC	HRS LOPN	1998	HHID FSUBHH OPN	49,013	4	LOPN ID
LX98_MC	HRS LOPN	1998	HHID FSUBHH OPN	4,031	4	LOPN ID; exit
LC00_MC	HRS LOPN	2000	HHID GSUBHH OPN	25,183	4	LOPN ID; LOPN has duplicate entries for 8 cases (removed)
LC00_MC_orig	HRS LOPN	2000	HHID GSUBHH OPN	46,023	4	LOPN ID; contains obs not in hrs9202
LX00_MC	HRS LOPN	2000	HHID RSUBHH OPN	4,273	4	LOPN ID; exit
LC02_MC	HRS LOPN	2002	HHID HSUBHH OPN	37,952	4	LOPN ID; LOPN has duplicate entries for 16 cases (removed)
LC02_MC_orig	HRS LOPN	2002	HHID HSUBHH OPN	70,126	4	LOPN ID; contains obs not in hrs9202

Table 15: (cont'd.)

filename	distribution	wave	ID	Nobs	Nvar	content & remarks
LX02_MC	HRS LOPN	2002	HHID SSUBHH OPN	8,297	4	LOPN ID; exit
kids	HRS CORE	1992	HHID ASUBHH OPN	24,697	48	children characteristics; transfers; 245 obs not in LOPN tracker
w2kids	HRS CORE	1994	HHID CSUBHH OPN	22,861	43	children characteristics; transfers; includes 120 exit interviews
h96pr_mc	HRS CORE	1996	HHID ESUBHH OPN	25,171	25	children characteristics;
h98pr_mc	HRS CORE	1998	HHID FSUBHH OPN	49,013	35	nonrespondent characteristics;
h00pr_mc	HRS CORE	2000	HHID GSUBHH OPN	46,023	33	nonrespondent characteristics;
h02pr_mc	HRS CORE	2002	HHID HSUBHH OPN	70,127	27	nonrespondent characteristics;
health	HRS CORE	1992	HHID PN	12,652	791	respondent health
househld	HRS CORE	1992	HHID ASUBHH	7,607	1,032	household wealth and income
w2b	HRS CORE	1994	HHID PN	11,596	183	respondent health
h96b_r	HRS CORE	1996	HHID PN	10,964	186	respondent health
h96i_jh	HRS CORE	1996	HHID ESUBHH	6,816	484	household wealth and income
h98b_r	HRS CORE	1998	HHID PN	21,384	190	respondent health
h98i_jh	HRS CORE	1998	HHID FSUBHH	14,395	416	household wealth and income
h00b_r	HRS CORE	2000	HHID PN	19,579	190	respondent health
h00i_jh	HRS CORE	2000	HHID GSUBHH	13,214	416	household wealth and income
h02c_r	HRS CORE	2002	HHID PN	18,166	195	respondent health
h02i_qh	HRS CORE	2002	HHID HSUBHH	12,350	424	household wealth and income
h96d_h	HRS CORE	1996	HHID ESUBHH	6,816	492	intrafamily relationships
h96d_tc	HRS CORE	1996	HHID ESUBHH OPN ETYPTRAN	3,901	22	transfer to children
h98d_h	HRS CORE	1998	HHID FSUBHH	14,395	523	intrafamily relationships
h98d_tc	HRS CORE	1998	HHID FSUBHH OPN FTRANNUM	6,176	27	transfer to children
h00d_h	HRS CORE	2000	HHID GSUBHH	13,214	545	intrafamily relationships
h00d_tc	HRS CORE	2000	HHID GSUBHH OPN GTRANNUM	5,839	27	transfer to children
h02e_h	HRS CORE	2002	HHID HSUBHH	12,350	108	intrafamily relationships
h02e_tc	HRS CORE	2002	HHID HSUBHH OPN HTC_NDX	5,133	24	transfer to children
h96d_mc_rev	HRS CORE	1996	HHID ESUBHH OPN	25,171	41	children-level information
h98d_mc_rev	HRS CORE	1998	HHID FSUBHH OPN	25,152	49	children-level information
h00d_mc_rev	HRS CORE	2000	HHID GSUBHH OPN	24,195	49	children-level information
h02e_mc_rev	HRS CORE	2002	HHID HSUBHH OPN	35,413	35	children-level information
h94i_n_rev	HRS IMP	1994	HHID CSUBHH	6,979	301	household-level income
h94i_k_rev	HRS IMP	1994	HHID CSUBHH	6,979	82	household-level wealth
h94i_kid_rev	HRS IMP	1994	HHID CSUBHH OPN	22,861	24	children-level information
h96i_dtc	HRS IMP	1996	HHID ESUBHH OPN ETYPTRAN	3,901	11	transfers to children (imputed)
h98i_dtc	HRS IMP	1998	HHID FSUBHH OPN FTRANNUM	6,176	11	
h00i_dtc	HRS IMP	2000	HHID GSUBHH OPN GTRANNUM	5,839	11	
h02i_etc	HRS IMP	2002	HHID HSUBHH OPN HTC_NDX	5,133	9	
rndhrs1e7	RANDHRS	1992	hhidpn	26,728	794	respondent wealth, health, work & income
rndhrs2e7	RANDHRS	1994	hhidpn	26,728	903	
rndhrs3e7	RANDHRS	1996	hhidpn	26,728	866	
rndhrs4e7	RANDHRS	1998	hhidpn	26,728	866	
rndhrs5e7	RANDHRS	2000	hhidpn	26,728	868	
rndhrs6e7	RANDHRS	2002	hhidpn	26,728	879	

2.4 Content of Final Data File

```
Contains data from all.dta
  obs:       136,797
 vars:           180                          13 Jul 2007 22:11
 size:    45,143,010 (64.1% of memory free)   (_dta has notes)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
HHID            str6   %9s                    HH identifier
LOPN            str4   %9s                    LOPN identifier
FAMSTORY        byte   %9.0g                  FAMSTORY identifier
STORY           str2   %9s
OPN             str3   %9s                    other person number ID
SUBHH           str1   %9s                    subhh ID
subhh           str1   %9s                    subhh ID
wave            byte   %8.0g                  wave
SUBHHIW         str6   %9s
H_WGTHH         int    %12.0g                 1992 Weight: Household-Level
FIN_WGTR        int    %9.0g                  FIN: R-level weight
FAM_WGTR        int    %9.0g                  FAM: R-level weight
PN_FAM          str3   %9s                    PN of wave-spec FAM R
PN_FIN          str3   %9s                    PN of wave-spec FIN R
HHFAMPN         str3   %9s                    PN HH FAM R (ignoring SUBHH)
HHFINPN         str3   %9s                    PN HH FIN R (ignoring SUBHH)
PNparent1       str3   %9s                    PN parent 1
PNparent2       str3   %9s                    PN parent 2
PNmother        str3   %9s                    PN mother
PNfather        str3   %9s                    PN father
intact1         str5   %9s                    no chg couple/sgle, may die
                                                (per wave)
intact2         str5   %9s                    no chg couple/sgle, noone may
                                                die (per wave)
intact3         str5   %9s                    no chg SUBHH between waves (per
                                                wave)
Kintact2        str5   %9s
Kdied           byte   %8.0g                  1992,imp: child reported
                                                dead[1]/alive[0]
KdiedF          byte   %9.0g
ACKintact1      byte   %8.0g                  no split'g into sub-HH, no chg
                                                in kids betw waves AC
CEKintact1      byte   %8.0g                  no split'g into sub-HH, no chg
                                                in kids betw waves CE
EFKintact1      byte   %8.0g                  no split'g into sub-HH, no chg
                                                in kids betw waves EF
FGKintact1      byte   %8.0g                  no split'g into sub-HH, no chg
                                                in kids betw waves FG
GHKintact1      byte   %8.0g                  no split'g into sub-HH, no chg
```

			in kids betw waves GH
ACKintact2	byte	%8.0g	may split into sub-HH, no chg in kids betw waves AC
CEKintact2	byte	%8.0g	may split into sub-HH, no chg in kids betw waves CE
EFKintact2	byte	%8.0g	may split into sub-HH, no chg in kids betw waves EF
FGKintact2	byte	%8.0g	may split into sub-HH, no chg in kids betw waves FG
GHKintact2	byte	%8.0g	may split into sub-HH, no chg in kids betw waves GH
ACint1	byte	%8.0g	no chg couple/sgle, may die, betw waves AC
CEint1	byte	%8.0g	no chg couple/sgle, may die, betw waves CE
EFint1	byte	%8.0g	no chg couple/sgle, may die, betw waves EF
FGint1	byte	%8.0g	no chg couple/sgle, may die, betw waves FG
GHint1	byte	%8.0g	no chg couple/sgle, may die, betw waves GH
ACint2	byte	%8.0g	no chg couple/sgle, noone may die, betw waves AC
CEint2	byte	%8.0g	no chg couple/sgle, noone may die, betw waves CE
EFint2	byte	%8.0g	no chg couple/sgle, noone may die, betw waves EF
FGint2	byte	%8.0g	no chg couple/sgle, noone may die, betw waves FG
GHint2	byte	%8.0g	no chg couple/sgle, noone may die, betw waves GH
ACint3	byte	%8.0g	no chg SUBHH between waves AC
CEint3	byte	%8.0g	no chg SUBHH between waves CE
EFint3	byte	%8.0g	no chg SUBHH between waves EF
FGint3	byte	%8.0g	no chg SUBHH between waves FG
GHint3	byte	%8.0g	no chg SUBHH between waves GH
Kamount1	float	%9.0g	Kid: transfer received 1992$ 1st impute
Kamount2	float	%9.0g	Kid: transfer received 1992$ 1st impute
Kamount3	float	%9.0g	Kid: transfer received 1992$ 1st impute
Kamount4	float	%9.0g	Kid: transfer received 1992$ 1st impute
Kamount5	float	%9.0g	Kid: transfer received 1992$ 1st impute
Kiamount	byte	%8.0g	Kid: transfer received, imp indic
KamtHRS	float	%9.0g	transfers made to child, HRS imputed value

KiamtHRS	byte	%48.0g	HRSIMP	imputation indicator: transfers to child HRS
KamountorgI	float	%9.0g		transfers made to indiv. child, original value
KamountorgA	float	%9.0g		transfers made to all children, original value
KbrackvI	byte	%9.0g		bracket indicator ind children
KbrackvA	byte	%9.0g		bracket indicator all children
KbrkvHRSI	int	%9.0g		HRS bracket indicator ind children, W8024S
KbrkvHRSA	int	%9.0g		HRS bracket indicator all children, E1449S
Ksexmale	byte	%8.0g		Kid: male
KIMsex	byte	%8.0g		Kid: missing: male
Kyob	int	%8.0g		Kid: year of birth
Kage	byte	%9.0g		Kid: currnt age
KIMage	byte	%8.0g		Kid: missing: currnt age
Kkids	byte	%8.0g		Kid: have kids
KIMkid	byte	%8.0g		Kid: missing: have kids
Kedu	byte	%8.0g		Kid: educ in yrs
KIMedu	byte	%8.0g		Kid: missing: educ in yrs
Knatural	byte	%8.0g		Kid: own child of (both) parnt(s)
KIMnat	byte	%8.0g		Kid: missing: own child
Kschool	byte	%8.0g		Kid: at school
KIMsch	byte	%8.0g		Kid: missing: at school
Kathome	byte	%8.0g		Kid: at home
KIMhom	byte	%8.0g		Kid: missing: at home
K10miles	byte	%8.0g		Kid: live <= 10 miles
KIM10m	byte	%8.0g		Kid: missing: live <= 10 miles
Khomeown	byte	%8.0g		Kid: home owner
KIMown	byte	%8.0g		Kid: missing: home owner
Kwrk	byte	%8.0g		Kid: works (1->30,2-<30,3-none)
Kwrkge30	byte	%8.0g		Kid: work >= 30 hrs/wk
Kwrklt30	byte	%8.0g		Kid: work < 30 hrs/wk
KIMwrk	byte	%8.0g		Kid: missing: works
Kmstat	byte	%8.0g		Kid: married (not: partnered)
KIMmst	byte	%8.0g		Kid: missing: married
Kinc1	double	%10.0g		Kid: lower income bracket
KIMinc1	byte	%8.0g		Kid: missing: lwr inc brckt
Kinc2	double	%10.0g		Kid: upper income bracket
KIMinc2	byte	%8.0g		Kid: missing: upr inc brckt
KBRinc	long	%20.0g	KBRinc	Kid: income bracket indic
KIMBRi	byte	%8.0g		Kid: missing: inc br ind
Kraceblck	byte	%8.0g		Kid: black (from parents)
Kracewhit	byte	%8.0g		Kid: white (from parents)
Kraceothr	byte	%8.0g		Kid: other non-white (from parents)
Khispanic	byte	%8.0g		Kid: hispanic (from parents)
H_COUPLE	byte	%8.0g		HH: couple indicator

H_ASSETS	double	%12.0g	HH: total assets (HRS)
H_HHINC	double	%9.0g	HH: total income (HRS)
H_rndasst	double	%10.0g	HH: total assets (RAND)
H_rndwlth	double	%10.0g	HH: total wealth, except IRA's (RAND)
H_rndinc	double	%12.0g	HH: total income (RAND)
FAM_ACmove	int	%9.0g	FAMR: miles moved betw waves AC
FAM_AGE	byte	%9.0g	FAMR: age
FAM_BIRTHYR	int	%9.0g	FAMR: year of birth
FAM_CEmove	int	%9.0g	FAMR: miles moved betw waves CE
FAM_deg2YC	byte	%9.0g	FAMR: highest degree: 2yr collg
FAM_deg4YC	byte	%9.0g	FAMR: highest degree: 4yr collg
FAM_degGED	byte	%9.0g	FAMR: highest degree: GED
FAM_degHS	byte	%9.0g	FAMR: highest degree: high sch
FAM_degMA	byte	%9.0g	FAMR: highest degree: MA
FAM_degNONE	byte	%9.0g	FAMR: highest degree: none
FAM_degPHD	byte	%9.0g	FAMR: highest degree: PhD
FAM_EFmove	int	%9.0g	FAMR: miles moved betw waves EF
FAM_FGmove	int	%9.0g	FAMR: miles moved betw waves FG
FAM_HEALTH	byte	%9.0g	FAMR: subj health
FAM_HISPANIC	byte	%9.0g	FAMR: ethn: hispanic
FAM_HTHCHG	byte	%9.0g	FAMR: change in health
FAM_hthexcel	byte	%9.0g	FAMR: health in 1992: excellent
FAM_hthfair	byte	%9.0g	FAMR: health in 1992: fair
FAM_hthgood	byte	%9.0g	FAMR: health in 1992: good
FAM_hthpoor	byte	%9.0g	FAMR: health in 1992: poor
FAM_hthvgood	byte	%9.0g	FAMR: health in 1992: very good
FAM_IMMI0054	byte	%9.0g	FAMR: immigrated before 1955
FAM_IMMI5564	byte	%9.0g	FAMR: immigrated 1955-1964
FAM_IMMI6574	byte	%9.0g	FAMR: immigrated 1965-1974
FAM_IMMI7599	byte	%9.0g	FAMR: immigrated after 1975
FAM_IMNEVER	byte	%9.0g	FAMR: never immigrated
FAM_RACEBLCK	byte	%9.0g	FAMR: black
FAM_RACEOTHR	byte	%9.0g	FAMR: other non-white
FAM_RACEWHIT	byte	%9.0g	FAMR: white
FAM_Region	byte	%9.0g	FAMR: region of residence
FAM_RegionB	byte	%9.0g	FAMR: region where born
FAM_RegLiv10	byte	%9.0g	FAMR: region lvg at age 10
FAM_SCHLYRS	byte	%9.0g	FAMR: yrs of schooling
FAM_SEXMALE	byte	%9.0g	FAMR: male
FAM_UrbRur	byte	%9.0g	FAMR: 1-Urb, 2-Sub, 3-Rural
FAM_USBORN	byte	%9.0g	FAMR: Born in US
FIN_ACmove	int	%9.0g	FINR: miles moved betw waves AC
FIN_AGE	byte	%9.0g	FINR: age
FIN_BIRTHYR	int	%9.0g	FINR: year of birth
FIN_CEmove	int	%9.0g	FINR: miles moved betw waves CE
FIN_deg2YC	byte	%9.0g	FINR: highest degree: 2yr collg
FIN_deg4YC	byte	%9.0g	FINR: highest degree: 4yr collg
FIN_degGED	byte	%9.0g	FINR: highest degree: GED
FIN_degHS	byte	%9.0g	FINR: highest degree: high sch

```
FIN_degMA       byte   %9.0g                 FINR: highest degree: MA
FIN_degNONE     byte   %9.0g                 FINR: highest degree: none
FIN_degPHD      byte   %9.0g                 FINR: highest degree: PhD
FIN_EFmove      int    %9.0g                 FINR: miles moved betw waves EF
FIN_FGmove      int    %9.0g                 FINR: miles moved betw waves FG
FIN_HEALTH      byte   %9.0g                 FINR: subj health
FIN_HISPANIC    byte   %9.0g                 FINR: ethn: hispanic
FIN_HTHCHG      byte   %9.0g                 FINR: change in health
FIN_hthexcel    byte   %9.0g                 FINR: health in 1992: excellent
FIN_hthfair     byte   %9.0g                 FINR: health in 1992: fair
FIN_hthgood     byte   %9.0g                 FINR: health in 1992: good
FIN_hthpoor     byte   %9.0g                 FINR: health in 1992: poor
FIN_hthvgood    byte   %9.0g                 FINR: health in 1992: very good
FIN_IMMI0054    byte   %9.0g                 FINR: immigrated before 1955
FIN_IMMI5564    byte   %9.0g                 FINR: immigrated 1955-1964
FIN_IMMI6574    byte   %9.0g                 FINR: immigrated 1965-1974
FIN_IMMI7599    byte   %9.0g                 FINR: immigrated after 1975
FIN_IMNEVER     byte   %9.0g                 FINR: never immigrated
FIN_RACEBLCK    byte   %9.0g                 FINR: black
FIN_RACEOTHR    byte   %9.0g                 FINR: other non-white
FIN_RACEWHIT    byte   %9.0g                 FINR: white
FIN_Region      byte   %9.0g                 FINR: region of residence
FIN_RegionB     byte   %9.0g                 FINR: region where born
FIN_RegLiv10    byte   %9.0g                 FINR: region lvg at age 10
FIN_SCHLYRS     byte   %9.0g                 FINR: yrs of schooling
FIN_SEXMALE     byte   %9.0g                 FINR: male
FIN_UrbRur      byte   %9.0g                 FINR: 1-Urb, 2-Sub, 3-Rural
FIN_USBORN      byte   %9.0g                 FINR: Born in US
KamntL          float  %9.0g                 transfers: lower bracket value
                                               (1991 USD)
KamntH          float  %9.0g                 transfers: upper bracket value
                                               (1991 USD)
-------------------------------------------------------------------------------
Sorted by:  HHID  LOPN  wave  FAMSTORY
```

2.5 Final Removal of Observations

The data set `all.dta` can be used for analysis. Note, however, that due to the imputations done at various levels, two additional changes should be applied:

1. there are multiple 'stories' in the data and presumably one should be selected for each child; for instance, selecting `FAMSTORY=1` ensures that only the first of all stories is taken into account—resulting in a loss of 7,737 (out of 136,797)

2. there are imputed values for waves where no interview was obtained from the household; these imputations were necessitated by controlling for the time path of evolution of some time varying variables (such as education), but may perhaps be cleaned out by conditioning on interview status; the following STATA code will do this, resulting in a data set with a final number of 102,827 observations

```
foreach W of num 1/6 {;
2.         g str1 interviewed=substr(SUBHHIW,'W',1);
```

```
3.         destring interviewed,replace;
4.         drop if interviewed!=1;
5.         drop interviewed;
6. };
```

2.6 Miscellaneous other files

- `dataalert.do`: takes care of the data alerts as spelled out on the HRS home page; last touched: Nov 11, 2005
- `MakeFamilyFromEarlyExit.do`: based on documentation in MakeFamilyFromEarlyExit.pdf; file is now obsolete as the Early Exit files in 1998 and 2000 have been replaced with final release versions.

3 Children: Permanent Income Regression

File: `Kperminc4.do`.

Our measure of children's permanent income is based on a random effects model that regresses current log household income (in 1991 dollars) on a number of observables. We include as regressors linear splines in age and education, an age/education interaction, sex, information on hours worked (below 30 hours per week, or 30 hours or more), and being married. Also, time (wave) dummies are included. We follow the methodology set out in Kapteyn *et al.* (2005) which calculates a time-varying measure of permanent income from a regression of log current income on observables. Unlike these authors, we do not take into account cohort effects.

The measure of current household income differs across survey waves. In the 1992 wave, only qualitative information is available, that is, we know if the child's annual (nominal) family income fell short of 10,000 dollars, exceeded 25,000 dollars, or fell in between. In the 1994 wave, parents are actually requested to supply an estimate of the amount of children's incomes, and if they were unable to do this, they were presented with a range card and asked to indicate an appropriate bracket with threshold values of 10,000, 25,000, and 40,000 dollars, respectively. In 1996 and in subsequent waves, child household income was elicited in a similar manner, except that the income thresholds were 10,000, 35,000, 50,000, and 100,000, and that brackets were elicited subsequently (unfolding bracket technique).

The regression model used takes this heterogeneous information into account in that it allows for continuous, discrete, and bracketed values in the endogenous variable; hence, the model is a generalized censored regression model as in Table **??**. It includes a composite error that has an individual specific random effect and an idiosyncratic error.

The model was estimated in the same way and with the same software as the two-level model for transfers, again using `FORTRAN` code. We add an estimate of the random effect to our predicted current income before calculating permanent income. This estimate is also known as 'empirical Bayes prediction'.

Estimation results for the auxiliary model are available on request. Having obtained the linear prediction, we convert log income to levels, and calculate permanent income by assuming a working life span ranging from 18 to 65. We discount future incomes at 4% per year. The resulting estimate of (annualized) permanent income is then obtained, and we discard a handful of observations with negative values and those with permanent income of more than $2m annual.

References

Kapteyn, A., R. Alessie, and A. Lusardi (2005): "Explaining the Wealth Holdings of Different Cohorts: Productivity Growth and Social Security," *European Economic Review*, 49, 1361-1391.