PSID A national study of socioeconomics and health over lifetimes and across generations

Technical Series Paper #08-05

Panel Study of Income Dynamics Revised Longitudinal Weights 1993-2005

Elena Gouskova, Steven Heeringa, Katherine McGonagle,
Robert Schoeni, Frank Stafford
Survey Research Center - Institute for Social Research
University of Michigan

May 2008

This project was supported by funding from the National Science Foundation (SES 0518943).

Technical Report

Panel Study of Income Dynamics

Revised Longitudinal Weights 1993-2005

May, 2008

Elena Gouskova, Steven G. Heeringa, Kate McGonagle, Robert Schoeni, Frank Stafford
ISR, University of Michigan, Ann Arbor, MI

1. Introduction

This document describes the construction of a revised set of 1993-2003 Core/Immigrant individual and family longitudinal sample weights for the Panel Study of Income Dynamics (PSID). These weights are a revised version of the longitudinal weights currently included in the PSID data public use data sets for these years. The new approach described in this document was also used to construct the 2005 core/immigrant longitudinal weights. The principal difference between the old and new weights is found in the approach for assigning weights to "reappearers", i.e., persons who return to the study after skipping one or more waves.

PSID analysts who use the revised longitudinal weights for cross-sectional analysis of a single wave of data will find increased case counts of sample persons with positive weights. Researchers conducting longitudinal analyses such as studies of simple change over time, general linear mixed modeling (e.g. growth curves), latent variable modeling or event history analyses will also have increased case counts; however, they will need to pay special attention to the pattern of missing data over time since the revised weights assign non-zero weights to a significant number of cases with incomplete data series for the period 1985-1993.

Prior to 1993, the number of PSID sample persons who "reappeared" in the interviewed sample after one or more waves of nonresponse was very limited—typically around 100-150 per year. Individual weight values for these returning cases were set to the weight they were assigned at the last wave in which they had participated. If the "reappearer" was a nonrespondent in a wave at which PSID weights were adjusted for cumulative panel attrition (i.e., 1969, 1974, 1979, 1984, and 1989), this procedure for carrying forward of the last recorded weight could introduce a small bias in the weighted population distribution for the PSID respondent sample. With such small numbers of prior wave nonrespondents "reappearing" in each wave, the potential size of any bias that resulted from this practical approach to creating a current individual weight for responding sample persons was very small.

In 1992, 1993 and 1994, the PSID was funded to conduct a major follow-up of sample persons who had been lost to nonresponse over the preceding decade. The nonresponse follow-up was very successful. From 1992 to 1994, PSID interviewed approximately 1800 persons who had not been interviewed since 1988. The recovery of former nonrespondents in 1994 was even more successful, with an additional 2900 persons who had been nonrespondents since 1988.

The success of the 1993 and 1994 recontact project presented a difficult choice in the creation of the 1993 and 1994 longitudinal weights. One option was to follow the conventional procedure of carrying forward the last available weight (the "reference weight") for "reappearers"/"recontacts". With appropriate updates to the correction for panel attrition, this option would maximize the case count of sample persons for cross-sectional analysis of each wave of PSID data. However, a longitudinal weight that assigned a population weight to "reappearers"/"recontacts" with extended sequences of missing data would be problematic for analysts who wished to conduct true longitudinal analysis. For example, an analyst interested in measuring simple change between 1992 and 1993 would choose the 1993 longitudinal weight (the terminal point weight). However, the large numbers of positively weighted "reappearers"/"recontacts" would have no data for 1992. If default case-wise deletion was permitted in the analysis, the weighted estimates of simple change would very likely be biased. Analysts with more advanced statistical background would have recognized this problem and possibly undertaken their own adjustments (propensity models, E-M type algorithms) to address the missing data problem but it could not be assumed that routine users of the PSID public use data would have access to the advanced statistical training required to apply these approaches. To ensure that the public use data sets would support robust inference in both cross-sectional and longitudinal analysis, a rule was introduced in the 1993 and 1994 weight development that assigned a zero individual weight to a sample person who had been a nonrespondent in 1989, which is the year of the last weight adjustment for panel attrition. The majority of sample persons who were assigned a value of zero for their individual weight under this rule had no data for the period 1989 to 1992—they had been PSID nonrespondents prior to 1989. So the decision was made to assign zero individual weights to those 1993 and 1994 reappearers/recontacts whose individual weight was not adjusted for attrition and mortality in 1989 and who, for the most part, had incomplete data for the 1989-1992 interval.

However, as the study continued into the 1990s, it became apparent that as a result of the 1993 and 1994 rule for handling the reappearers/recontacts, the number of sample persons with zero-valued individual weights was becoming increasingly large, and so was the number of families with zero family weights. This resulted in substantial losses of usable cases for PSID analysts interested in single-wave cross-sectional analyses for 1993 and following waves, or for short-term longitudinal analysis involving the post-1992 waves of data. To maximize the available case count for sample persons at each wave, it was decided to revise the longitudinal weights.

Effectively, the new weight variables contain a positive weight for each PSID sample person at each wave (1993-2005) in which they are associated with a responding family. The revised individual weights make no exclusions of sample persons based on their history of response/nonresponse in the PSID panel. Consequently, researchers interested in true longitudinal analysis are encouraged to carefully examine the patterns of missing data across waves of interest before proceeding with their longitudinal analysis. With the new, revised longitudinal weights, the sample size available for weighted analysis has increased; in 2003 the sample of individuals with positive weights has increased by about 30% and the sample of families has increased by 19%. In addition, the new weight assignment strategy makes it easy to identify sample persons with non-zero data, all of whom now carry a positive individual weight.

This document provides details on the construction of the revised series of longitudinal weights and is organized as follows. First, the PSID sample is described in Section 2. Details are provided on the PSID sample design, composition, following rules and recontact efforts. Section 3 provides background on the sample design and following rules with a focus on the period starting in 1990 when some changes were made to these rules. Details on the methodological approaches in construction of the revised 1993-2005 Core (Core/Immigrant) longitudinal weights are provided in Section 4. Section 5 describes the revised weights series, 1993-2005, and presents the resulting weights. Case counts and distributional statistics for the constructed weights are presented. Section 5 also includes a comparison of weighted distributions of some sample characteristics in the PSID with those of the Current Population Survey (CPS).

2. The PSID Sample

The PSID's dynamic sample design and following rules are the building blocks for the strategy used in weight construction, the assignment of weights, and the use of weights in different types of analysis. The following rules are important for understanding how the weights are constructed, who gets a positive weight and who does not, and how weights should be used in different types of analysis[1]. The next three subsections provide a brief overview of the PSID sample design and sample composition. These sections also describe the PSID sample following rules during the period 1989 through 2005.

2.1 The PSID sample design and composition

The PSID started in 1968 with 4,802 families. The sample was comprised of two separate samples: an equal probability national sample of households selected from the Survey Research Center 1960 National Sample (SRC) and a subsample of families interviewed in 1967 by the Bureau of the Census for the Office of Economic Opportunity (SEO). The SEO sample included poor families with income levels twice below the 1967 federal poverty line. The resulting combined SRC and SEO sample, referred to here as the Core sample, is an unequal selection probability sample. Compensatory weights were developed in 1968 to account for the differential sampling rates used to select the SEO and SRC components of the PSID. (A description of these weights may be found at the PSID web site (http://psidonline.isr.umich.edu/data/weights/).

[1] Weighting choice for different types of statistical analyses using PSID data is also discussed in *Heeringa et. al. (2008)*

The initial PSID sample is not representative of individuals who immigrated to the US after 1968. Two independent efforts have been made to address this limitation. First, in 1990 the PSID added 2,043 Latino households, representiong the three largest Latino groups in the country: Mexican-, Puerto Rican-, and Cuban-American-[2]. But while this sample did include some major groups of immigrants, it missed out on the full range of post-1968 immigrants, and those of Asian descent, in particular. Because of this crucial shortcoming, and a lack of sufficient funding, the Latino sample was dropped after 1995. Today, most analysts using the panel aspects of the PSID do not use the data from the Latino supplement.

The second effort to incorporate post-1968 immigrants was in survey years 1997 and 1999, when 511 in 1997 and 70 in 1999[3] immigrant families were added to the PSID. The sample was obtained by screening a representative sample to determine if the family was headed by an individual who was neither living in the U.S. in 1968 nor a US citizen or, if born since 1968, was the child of parents who were neither U.S. citizens nor living in the U.S. in 1968. If the head has a spouse or long-term cohabiter present in the family, then that person must also have met the criteria for immigrant status. In addition, the individual or spouse/long-term cohabiter had to have resided in the U.S. since January 1 of two calendar years prior. By 2005, the number of post-1968 immigrant families had increased to 572 because some family members had split off into their own family units.

1997 was also the year that the SEO sample was reduced to achieve cost savings. While many SEO families were eliminated, 43%, or 1,714 families, remained in the active sample. The reductions were achieved by dropping entire family trees, i.e., individuals with the same 1968 ID. This approach preserved the maximum number of family trees for support of intergenerational studies. Through natural sample growth generated by split-offs since 1997, the SEO sample had grown to include 2,279 families in 2005.

As of 2005 there were 8,041 PSID families. All of these families are members of the Core or the Immigrant samples, labels that help to distinguish them from the Latino sample families that were interviewed during 1990-1995 interview years. The

---

[2] For fuller description of the Latino sample refer to PSID documentation for 1990 interview year (http://psidonline.isr.umich.edu/Data/Documentation/pdf_doc/psid90w23.pdf).
[3] The 70 families were screened in 1997 but had not been in the U.S. continually for 2 years.

longitudinal weights described in the document pertain only to the Core and the Immigrant samples and will be referred as the Core longitudinal weights for 1993-1996 or the Core/Immigrant weights for 1997-2005 survey years.

2.2 PSID following strategies and the sample person concept

Throughout this section and the next we will describe the PSID following rules and recontact efforts, which are summarized in Table 1 for the period 1990 through 2005 during which some of the rules were changed.

After the first interview, PSID sample members, including all those leaving to establish separate family units, are tracked and followed. The PSID rules for following household members were designed, with weights, to maintain a nationally representative sample of families at any point in time as well as across time, absent immigration. As described above, a sample of immigrants arriving in the US after 1968 was added in 1997/1999.

Children born to or adopted by an original sample person are classified as sample members and are eligible for tracking as separate family units when they set up their own households[4]. These individuals, as well as any other family unit (FU) members who separate from the reinterview family to establish separate household, are referred to as "split-offs" and tracked to their new family units. This procedure replicates the population's family-building activity and produces a dynamic sample of families each

---

[4] A child who is born to or adopted by a parent who is a sample member becomes a sample person if the child was born after the initial interview for the specific sample. For SRC and Census samples, this means that the child was born after the 1968 interview; for Latinos, after the 1990 interview (sample status for Latino sample cases added in 1992 was based on considerations of who was living with the targeted original sample family in 1990); and for Immigrants, after the 1997 or 1999 interview. Unlike the Latinos in 1992, sample membership for those Immigrants who were added in 1999 was based on the 1999 family.

In an adoptive situation, the child must have no prior blood or marriage kinship ties to the sample parent. For example, a stepchild who is adopted by a sample member does not qualify for sample membership. In other words, the fact of an adoption by a blood or step relationship does not change a child's sample status, whereas adoption by a nonrelative does change it. A sample child who is adopted by unrelated foster parents becomes nonsample; a nonsample child who is adopted by an unrelated sample member becomes sample.

year. New PSID families form when children grow up and establish separate households or when marriage partners go separate ways.   With a high reinterview rate and a high success rate in adding split-offs, sample growth occurs over time in both the number of family units and the number of people residing with a sample member at some time during the study.

For understanding the PSID following strategies the concept of the *sample person* is crucial. Through 1993, a sample person was defined as someone who is either an original sample individual or an offspring born to or adopted by a sample individual who at the time was actively participating in the study; the child had to appear in the study more or less at birth. In 1994 this definition was expanded to include children born to or adopted by a sample person when the sample person was <u>not</u> participating in the study; the child need not have moved into a responding panel family at birth.

The main PSID following rule, called ”base” in Table 1, states that a sample person who responded in the previous wave is eligible to be followed. Through 1992, sample members under 18 years of age were never followed in their own right if they left the family unless: a) they had set up their own independent households or b) they moved to the new home with an adult sample member. In 1993 PSID relaxed its rules so that these younger persons were followed and an interview with an adult in the new family group was attempted. As a corollary to this alteration of a long-established tradition, the family composition rules changed. PSID families had always included a sample member as the Head or Wife/“Wife” of the family unit, but this became impossible in some cases where underage sample members moved with a nonsample parent. Therefore, both the Head and the Wife/“Wife” may be nonsample persons beginning 1993.

In two instances the PSID targeted and interviewed particular groups of nonsample respondents to support specific research areas. First, “nonsample elderly” persons who were 65 or older were followed between 1990 and 1995. Second, the “nonsample parents” of young sample children were followed starting in 1994, but this practice was ended for the 2005 and succeeding waves (see Table 1). As a result of this following of nonsample respondents, during the 1990-2003 period some PSID families may not include a sample member.

2.3 The recontact efforts

Prior to 1990, PSID did not attempt to contact persons who had become nonrespondents in previous waves. Some of these respondents, however, reappeared as respondents in a current wave's data collection through ties to cooperating families. The annual totals for these "reappearers" were relatively small for the period 1968-1990. In 1990, the PSID made the first attempt to locate and interview prior wave non-respondents. As part of the supplemental study of elderly, the PSID contacted about 200 sample and nonsample individuals older than 65 who had been nonresponse since 1985 ("elderly recontact" in Table 1).

In 1995, as part of another supplemental study, PSID attempted to contact nonsample parents of children under 18 who were nonresp prior to 1995 ('nonsample parent recontact" in Table 1).

In 1992, 1993 and 1994, the PSID was funded to conduct a major follow-up of sample persons who had been lost to nonresponse over the preceding decade ("primary recontact" in Table1). During these years about 500, 1000, and 2000 prior nonrespondents were reintroduced in the study. Among these recontacted persons the corresponding numbers of sample persons were about 400, 700 and 1500.

Starting in 1997, PSID also attempted to contact sample individuals who did not respond in the prior wave , but responded in the wave before the prior wave. This group of respondents is called "established recontact" (see Table 1).

3. Overview of the Longitudinal Weights and who gets a positive weight

The Core (Core/Immigrant) longitudinal weights are designed to enable unbiased estimation of the descriptive statistics for the U.S. individuals and families that were eligible for the PSID survey population.[5] Since the Latino sample is not included, the Latino sample members carry zero value for the Core (Core/Immigrant) longitudinal weight.

[5] As noted, before the 1997 wave the eligible families excluded those who immigrated into the U.S. after 1968. Addition of the Immigrant subsample in 1997 and 1999 allowed the post-1968 immigrant families in the pool of eligible families.

Sample persons

Under the revised Core/Immigrant longitudinal weight computation algorithm, a non-zero positive weight value is assigned to each cooperating PSID sample person. Each interviewed family that includes one or more sample persons receives a positive, non-zero family weight.

Nonsample persons

The PSID is designed to be a longitudinal study of individual and families. Each year, individual data are collected for nonsample persons who enter a PSID family through marriage or residency. Data for nonsample persons presents a problem for longitudinal analysis since the time series for these individuals is left censored at the date at which they entered the PSID family. Furthermore, it is not likely that this left censoring is random with respect to the types of variables that might be considered in longitudinal analysis.

Because of the left censoring of their data series, nonsample persons in PSID families have historically been assigned a zero value selection weight factor and zero value for the PSID longitudinal weights. The 1993-2005 longitudinal weights continue this practice: all nonsample persons interviewed in 1993-2005 have zero longitudinal weight[6].

Reappearers and recontact persons

PSID assigns the last available positive weight, called the reference weight, to sample persons who appear in the study after being nonresponse in prior waves. A note is warranted with regard to this rule. Researchers should be aware of a possible bias that may arise when the number of reappearers (recontact persons) is large and their preceding years of nonresponse happen to include a year in which a nonresponse adjustment was performed[7]. The weights of their counterparts, those who responded continuously, were adjusted for the reappearers' nonresponse in previous waves.

---

[6] Note that the Core (Core/Immigrant) cross-sectional weights available for 1997 -2005 survey years provide a positive weight for all individuals in the PSID, both sample and nonsample members (see Heeringa et. al. (2008))

[7] Prior to 1993 the nonresponse adjustments were performed in 1969, 1974,1979, 1984,1989.

Although in years prior to 1993 the total annual number of reappearers was relatively small with a small associated risk of bias, a considerable number of sample persons was reintroduced into the study in 1993 and 1994. There is also a possibility of a bias in assuming that each sample person assigned a positive weight at time *t* has the requisite data to conduct a weighted longitudinal analysis for the multi-year interval (t-k,t). Depending on the time interval of interest, the reappearers/recontacts may not, in fact, have the required data for the previous year(s). Researchers who are conducting longitudinal analysis of the PSID data are encouraged to check how many cases are excluded from the analysis due to nonresponse for an entire wave or item missing data for key variables.

4. Methodological Approach to the Longitudinal Weight Construction

PSID longitudinal weights are inverse probability weights. In each year weight construction for the PSID family and individual weights follows a two-step procedure.

- In the first stage, individual longitudinal weights are constructed for all individuals who ever participated in the study.
- In the second stage, the family weight is calculated as the average of the individual weights of all the family members—both sample and nonsample --- participating in the study in that year[8].

Since derivation of family weights from the individual weights is straightforward, we focus on the construction approach used for the individual weights. There are two steps involved in the approach to the construction of the individual weights. First, the sample is divided into strata according to a person's status in two consecutive waves. Note that in the process of weight construction we always consider two waves, which we will refer to as "the prior year" and "the current year". In the following discussion, the current year is also denoted as t and the prior year as t_0. In the

[8] Prior to 1993, a family weight was calculated as the average of the head and wife/"wife" weights. The modification in family weight calculation was a result of the change in the following rule described in section 2.2. Starting in 1993following sample children regardless of age created a possibility of families with neither the head or the wife being a sample member.

second step, each person is given an individual weight depending on the stratum to which the person belongs.

The next two subsections provide details on these two steps. The last subsection discusses the derivation of the attrition adjustment factor used in the weight construction.

4.1 Step1: The sample division into strata

Each individual who has ever participated in the study is assigned to one of ten strata depending on his/her status transition between t_0 and t. These strata can be generalized into six main groups: 1) response, those who responded in the previous and the current waves; 2) recent exit, those who left the study between the previous and the current wave, 3) past exit, those who appeared in the study at some point in time but left it before the previous wave; 4) recent entry, those who first entered the study at the time of the current wave; 5) reentry/re-contact, those who participated in the study at some point in time, were nonresponse in the previous wave and voluntarily re-entered the study or responded to re-contact during the current wave; 6) future entry, those who have not yet appeared in the study as of the time of the current wave but will enter it in a later wave. Each group consists of one stratum, with the exception of two: the recent entry and recent exit. Recent entry and recent exit groups are each comprised of multiple strata. Table 2 reports numbers of persons in each stratum for selected pairs of the current and prior years.

Recent entry

Currently, a person has one of four main entry routes into the study. First is a child who meets the qualifications above and is born in to a responding sample family. This group also includes very young adoptees who are placed with the adoptive parents quite soon after birth ("Sample born in" in Table 2).

Second is a child who otherwise meets the tests above but is not born into a responding sample family. Such a child may be born to or adopted by a sample person who has left the panel while that sample person is nonresponse. Also included here are older children who are adopted into responding family units. Before 1994, children who had moved in subsequent to birth were not considered sample individuals and therefore were not given positive weights. In 1994, the definition of a sample member was changed to include such movers-in, and, thenceforth, these individuals received a positive weight ("Sample mover in" in Table 2).

The third way one can enter the panel is by marriage or through cohabitation with a sample person. Such an individual is a nonsample member and receives a zero weight ("Nonsample" in Table 2).

Finally, one can become a panel sample member by being part of a responding family unit in the wave in which a specific sample is initiated, e.g., the SRO or Census samples in 1968 or the samples of Immigrants in 1997 and 1999. Naturally, we have one additional twist to this rule: the "appearers". An appearer is a sample member who, through an error in the initial family listing, is subsequently discovered to have been present in a family unit since the inception of the sample. This situation arises very infrequently and usually happens within a wave or two of the original interview ("Sample new"/"Sample appear" in Table 2).

Recent exit

Recent exit is another complex group consisting of two substrata. Exit from the panel occurs because of mortality and nonresponse. The nonresponse group includes cases where a family or a sample person refuses to participate or cannot be located or contacted.

4.2 Step2: Assignment criteria for the individual weight

Next in the development of the individual weight is assignment and computation of weights. The individual weight assignment occurs in two phases. First, weights are assigned to individuals in all strata with the exception of the recent entry group. Then weights are assigned to those who entered the panel for the first time based on the individual weights of the other family members.

The weight assignment follows the same rule for individuals in the same strata. Individuals in groups that are nonresponse in the current-wave data, for reasons including past exit, future entry, and recent exit, get assigned a zero weight. The reentry/recontact group, i.e. those who were nonresponse in the previous wave, are given a weight equal to their last available non-zero weight, i.e., the reference weight. The weight assignment rule for those in the recent entry group varies by stratum. A new nonsample individual is given an individual weight equal to 0. Born in and mover in sample individual weights are equal to an average of head and wife/"wife" individual weights. The new sample individuals in the last stratum, new sample individuals/reappearers, receive a weight equal to the inverse of selection probability. If the case is a reappearer, then the weight is the average of individual weights among the sample members of the family.

Two approaches are generally used in PSID to construct individual weights for the response group. In both cases the current-wave weights are based on the prior period's individual weights. The difference is that in one case the adjustment is made to account for attrition and mortality between the waves, while in the other instance, no such adjustment is made. We refer to the weights obtained with the former approach as the adjusted weights and to the weights obtained with the latter approach as the carry-over weights. In the carry-over weights the individual weight in the current wave is taken to be the same as the weight in the previous wave. For adjusted weights, the individual weight is a product of the previous wave weight and the attrition adjustment factor. The details on the derivation of the factor are given in the next subsection.

Table 3 summarizes the weight assignment rules for adjusted weight and carry-over weights.

4.3 Attrition adjustment factor

This subsection provides theoretical justification for the way the attrition adjustment is made. It also describes the approach used to calculate the attrition adjustment parameter in the PSID. The discussion in this subsection considers only those who were respondents in the prior period and in the current period are either in the response group or in the recent exit group.

The main sources of attrition among the respondents who participated in the study in previous years include: nonresponse of living panel members; loss to follow-up where the death of the sample person is unknown; and known deaths to panel members. The table below shows the pattern of vital status among those who responded at t conditional on the interview outcome at t . The vital status is denoted by A, and interview status is denoted by S. Note that when the interview status is response, $S_t = 1$, or the interview status is known dead, $S_t = 2$, then the vital status is known. However, when the status is nonresponse, $S_t = 1$, we have no information whether the person is living or dead.

	Vital status	
Interview outcome	$A_t = 0$ (dead)	$A_t = 1$ (alive)
$S_t = 1$ (response)	0	X
$S_t = 2$ (known dead)	X	0
$S_t = 3$ (nonresponse: refusal or lost to follow up)	X (exact number is not known)	X (exact number is not known)

To understand our adjustments, it is useful to note that the probability of the intersection of S_t and A_t can be decomposed into a product of two other probabilities as follows:

$$\Pr(S_t, A_t) = \Pr(A_t) * \Pr(S_t \mid A_t)$$

When the purpose of the weight is to represent the *t* population, the conditional probability, $\Pr(S_t \mid A_t)$, should be the basis of the *t* weight, not $\Pr(S_t, A_t)$. Only those surviving to *t* are still in the population at that time, and we want to know just how many people each surviving person who responds represents in the population at time *t*. $\Pr(S_t \mid A_t)$ tells us what the probability of an individual responding in *t* was, given that person was indeed among those who survived until *t*. Thus, this is the probability that we want to incorporate into our weight adjustment. The reciprocal of the conditional probability, $\left[\Pr(S_t = 1 \mid A_t = 1)\right]^{-1}$, is the factor we use to adjust the weight for differential nonresponse and mortality.

Noting that the probability of response, $\Pr(S_t = 1)$, is the same as probability of response and survival, $\Pr(S_t = 1, A_t = 1)$, we get

$$\Pr(S_t = 1 \mid A_t = 1) = \frac{\Pr(S_t = 1, A_t = 1)}{\Pr(A_t = 1)} = \frac{\Pr(S_t = 1)}{\Pr(A_t = 1)}$$

Thus, conditional on being alive, the probability of responding is a ratio of the probability of response, $\Pr(S_t = 1),$ and the probability of being alive, $\Pr(A_t = 1)$.

The probability of being alive can be rewritten as:

$$\Pr(A_t = 1) = \sum_{j=1}^{3} \Pr(A_t = 1 \mid S_t = j) \cdot \Pr(S_t = j) =$$
$$\Pr(S_t = 1) + \Pr(A_t = 1 \mid S_t = 3) \Pr(S_t = 3)$$

The second line follows because $\Pr(A_t = 1 \mid S_t = 2) = 0$ and $\Pr(A_t = 1 \mid S_t = 1) = 1$. Denote $\Pr(S_t = j)$ as q_j, then the expression for the conditional probability becomes

$$\Pr(S_t = 1 \mid A_t = 1) = \frac{q_1}{q_1 + \Pr(A_t = 1 \mid S_t = 3) \cdot q_3} \qquad (1)$$

q_j, where $j \in \{1,2,3\}$, can be estimated directly from the data. However, the conditional probability of being alive, $\Pr(A_t = 1 \mid S_t = 3)$, is not possible to estimate from the data, and so one needs to have additional information or to make an assumption.

Prior to 1993 in the weight construction process, the numerator and denominator of the ratio (1) were estimated separately. The numerator was estimated from the data using the weighting cell approach. The denominator was obtained with the use of the national estimates of mortality rates. This approach was based on the assumption that mortality rates in the sample are the same as the mortality rates in the population. The national estimates of mortality rates given in Vital Statistics, U.S. Census Bureau, were used to construct age-gender-race-specific probabilities of survival, which served as estimates of $\Pr(A_t = 1)$.

From 1993 forward the conditional probability (1) was estimated simultaneously with the use of a statistical modeling approach. In this approach we assumed a value for

$\Pr(A_t = 1 | S_t = 3)$ from the [0,1] interval. In particular we assumed that the conditional probability is 1, which would mean that all nonrespondents are living. While this is a strong assumption, it appears to be a reasonable approximation in the PSID case. The same assumption is made in the analysis of attrition in Health and Retirement Study (HRS) by Kapteyn, Michaud, Smith, van Soest (2006).

The multinomial logistic model (Maddala, 1983) was used to estimate probabilities of three states $q_j (j = 1,2,3)$. The probability of being in state *j* by household *i* that responded in the previous wave is given by

$$P(S_{i,t} = j | x_{it_0}) = \frac{\exp(\beta_j ' x_{it_0})}{\sum_{j=1}^{3} \exp(\beta_k ' x_{it_0})}, \qquad (2)$$

where x_{it_0} is a vector of observations on a set of characteristics at the previous wave and β_j is a vector of coefficients. Identification requires that some constraint be imposed on the β vectors. We employ the following constraint:

$$\beta_1 = 0 \qquad (3)$$

The model is estimated on the subset comprised of the *response* and the *recent exit* groups, i.e., those for whom data were collected in the previous wave.

4.4 Scaling

The revised PSID family and individuals weights are scaled to arbitary total counts that were established when the initial weights were created for the 1968 Wave I probability sample. With the obvious exception of weighted estimation of population totals, statistics estimated from the PSID data should be invariant to any linear scaling (multiplication or division by a constant) of the family or individual weights. Analysts may wish to rescale the weight values to suit their individual preference for requirements of their selected software systems or programs.

One common procedure is to simply rescale the weights to the appropriate U.S. Census population counts/estimates or Current Population Survey (CPS) estimates of the size of the target population in the year of interest:

$$W_{i,pop} = W_i \cdot \frac{N_{pop}}{\sum_{i=1}^{n} W_i}$$

where:

$W_i =$ the revised PSID weight value for unit i (family, individual);

$N_{pop} =$ the U.S. Census or CPS value for the target population total;

$W_{i,pop} =$ the rescaled PSID "population weight", $\sum_{i=1}^{n} W_{i,pop} = N_{pop}$.

A second procedure commonly seen in public use data sets is to "normalize" the weights or rescale the weights so that they sum to the nominal sample size for the set of observations included in the analysis:

$$W_{i,norm} = W_i \cdot \frac{n}{\sum_{i=1}^{n} W_i}$$

where:

$W_i =$ the revised PSID weight value for unit i (family, individual);

$n =$ the count of PSID sample families or persons included in the analysis;

$W_{i,norm} =$ the rescaled PSID "normalized weight", $\sum_{i=1}^{n} W_{i,norm} = n$.

In past decades, the practice of normalizing the weights was often a necessity since statistical software systems employed the case weights as frequency counts of observations (e.g., a weight of 100 was treated as 100 duplicate cases) resulting in biased estimates of variances, covariances and pseudo likelihood functions that are required in the computation of standard errors and inferential statistics for simple statistics such as means or more complex statistics such as multivariate regression coefficients. In today's statistical software systems, it should not be necessary for researchers to normalize weight values prior to conducting analysis; however, a few of the older programs such as

SAS Proc Logistic still require the user to specify an option to normalize the weights prior to analysis. Analysts are encouraged to review the documentation for their chosen software system to determine how weights are applied in the programs that they will use for analysis of PSID data.

5 The Core (Core/ Immigrant) Longitudinal Weights, 1993 through 2005

5.1 Description

The specific methods used in the construction of the Core (Core/Immigrant) longitudinal 1993 through 2005 weights in each year are summarized in Table 4. The 1993 longitudinal weights are attrition-adjusted weights constructed to account for losses between 1989 and 1993. Table 5 reports the result of the multinomial logistic regression. The weights in the following three years, 1994, 1995 and 1996, are carry-over weights.

Year 1997 is marked by the reduction of a portion of the SEO sample and by the addition the Immigrant sample. To account for the partial reduction of the Core sample, a weighting cell process was used. All PSID Core individuals were assigned to one of the 24 cells based on age (age<=25, 25<age<=40, 40<age<=60, age>60), gender (male, female) and race of head (white, black, other). For each cell, the weights of sample individuals who responded in 1997 were increased to compensate for loss of individuals who responded in 1996 but did not in 1997.

1997 PSID Immigrant Supplement sample families were initially selected with equal probability. During the field period, sample replicates of area segments with higher expected prevalence of immigrant households were oversampled to increase the efficiency of the household screening. Each of the interviewed families in the 1997 Immigrant Supplement was assigned an initial base weight value that reflected the probability of selection and screening for the area segment in which they resided.

To combine the weights from the Core and the Immigrant samples, the Immigrant sample weights were scaled so that the proportion of the Immigrant sample in the combined sample is 7%, which is an estimate of the proportion of the households that have immigrated to the United States since 1968[9].

---

[9] Based on data from the 1997 CPS, an estimated 7.5 percent of U.S. households have immigrated to the United States after 1968.

Similar to 1997, the 1999 weights are constructed separately for the Core and the Immigrant samples and then combined using the 93:7 ratio. The Core sample weights are carry-over weights. In 2001 the Immigrant and Core samples are treated the same way, with weights for both being carried over from the previous wave.

The 2003 weights are attrition-adjusted weights. They account for the attrition between 1999 and 2003 among 1999 study participants from the Core and the Immigrant samples. Table 6 reports the result of the multinomial logistic regression.

Finally, the 2005 weights are constructed by using the carryover approach summarized in Table 3.

The resulting longitudinal weights are stored in the PSID data archive under names provided in Table 15.

5.2 Estimates

For each PSID wave from 1989 through 2005, Tables 7 and 8 report the total number of cases where the revised longitudinal weights are positive, zero, or missing for individuals and families, respectively. Table 7 also provides information on the total number of individuals in the study and the number of sample and nonsample persons. Table 8 shows the total number of families in the study each year, and how many families have no sample person. Information is given separately for the Core/Immigrant sample and the Latino sample.

Table 7 shows that, starting in 1993, all sample persons from the Core/Immigrant sample are assigned a positive individual longitudinal weight and nonsample persons are assigned a zero weight. Members of the Latino sample families are also assigned a zero individual longitudinal weight.

Table 8 shows that the 1993-2005 Core/Immigrant longitudinal family weight can be zero in only two instances: when the family belongs to the Latino sample or when the family has no sample person associated with it. (Recall from the discussion above that prior to 2005 the following rules were such that some families could contain no sample members.) In all other cases the family weight is a positive number.[10]

---

[10] Note that for a handful of cases families have no sample members but they have positive weights. This most likely arose because there were individuals in these families who at the time of interview were

Tables 9 and 10 report summary statistics for the longitudinal individual and family weights, respectively. The most important pattern is the increase in the mean and variance in 1997 resulting from the elimination of some of the SEO sample and addition of the immigrant sample.

Tables 11 through 15 compare distributions of selected characteristics, including age, gender, race, and family income in the PSID data obtained with and without the longitudinal weights to those in the Current Population Survey (CPS) between survey years 1990 and 2005. These tables are useful for examining three features of the PSID data: consistency of unweighted and weighted estimates across years, effect of the longitudinal weights on the distributions of the characteristics, and, finally, the closeness of the PSID estimates with those obtained with the CPS data[11].

The tables show that consistency across years of weighted distributions is comparable to the consistency of unweighted distributions. An exception is for race and income between 1996 and 1997; the changes in the sample (i.e., elimination of part of the SEO sample and addition of the immigrant sample) led to a one-time change between 1996 and 1997 in the unweighted racial and income distributions. Comparison of the unweighted and weighted PSID distributions with the CPS distributions reveals that weighted estimates in a majority of cases are closer to CPS estimates than are the estimates obtained without weights. In addition, the PSID and CPS align fairly closely in most cases. An exception is for the racial distribution, and this arises for at least two reasons. Prior to 1997 the PSID sample did not include post-1968 immigrants, the majority of whom would be classified as white. In addition, since PSID weights were never benchmarked to the CPS or another federal source such as a post-1968 Census, we believe that the original 1968 computations for the SEO/SRC joint probabilities and

---

believed to be sample members, but at some later wave were redefined as nonsample based on additional information.

[11] Note, that some characteristics are not strictly comparable between the two surveys. For example, in the PSID race is not asked for everybody while in the CPS it is. To calculate proportions of black and non-black individuals in the PSID data, individual race was approximated with the race of the family head. Second, measures of educational attainment differ between the surveys because of how these questions are formulated and because of changes in the question sequences that occurred over the years. Finally, the family income measure in the PSID is not directly comparable to the CPS household income measure due to the difference in definition of family unit in PSID and household unit in CPS.

subsequent attrition adjustments in 1969, 1974, 1979, 1984 and 1989 may have led to a small positive bias in the weighted representation for black families and individuals.

References

Heeringa S., E. Gouskova and K. Mcgonagle, 2008. " Cross-sectional Analysis of PSID data and Cross-sectional weights 1997-2005", technical report, ISR, University of Michigan.

Kapteyn A., P. Michaud,J. Smith and A. van Soest, 2006."Effects of Attrition and Non-Response in the Health and Retirement Study," IZA Discussion Papers 2246, Institute for the Study of Labor (IZA).

Maddala, G.S. 1983. Limited Dependent and Qualitative Variables Econometrics. Cambridge: Cambridge University Press

Table 1: Following Strategies

		1990	1991	1992	1993	1994	1995	1996	1997	1999	2001	2003	2005
Who is followed	base	S, R in 1989, 18 or older	S, R in 1990, 18 or older	S, R in 1991, 18 or older	S, R in 1992	S, R in 1993	S, R in 1994	S, R in 1995	S, R in 1996	S, R in 1997	S, R in 1999	S, R in 2001	S, R in 2003
	non-sample parent					NS parent, R in 1993 , has a S child who is 18 or younger	NS parent, R in 1994 , has a S child who is 18 or younger	NS parent, R in 1995 , has a S child who is 18 or younger	NS parent, R in 1996 , has a S child who is 18 or younger	NS parent, R in 1997 , has a S child who is 18 or younger	NS parent, R in 1999 , has a S child who is 25 or younger	NS parent, R in 2001 , has a S child who is 25 or younger	
	non-sample elderly	NS, R in 1989, 65 or older	NS, R in 1990, 65 or older	NS, R in 1991, 65 or older	NS, R in 1992, 65 or older	NS, R in 1993, 65 or older	NS, R in 1994, 65 or older						
Who is excluded from being followed								Latino sample is dropped	A part of SEO is dropped				
New sample added		Latino sample (2,043 families)		Latino sample (265 families)					Immigrants sample (441 families)	Immigrants sample (70 families)			
Recontact	established recontact								S,NR in 1996, R in 1995	S, NR in 1997, R in 1996	S, NR in 1999, R in 1997	S, NR in 2001, R in 1999	S, NR in 2003, R in 2001
	primary recontact			S, NR by 1990	S, NR by 1992	S, NR by 1992							
	non-sample parent recontact					NS, paren, has a S child who is 18 or younger							
	elderly recontact	S and NS, NR in 1985-1989,65 or older											
Notes:	1) S=sample, 2) NS=nonsample, 3) R=response, 4) NR=nonresponse												

Table 2: Number of individuals in each group/strata

Group	Stratum	t0=1989 t=1993	t0=1993 t=1994	t0=1994 t=1995	t0=1995 t=1996	t0=1996 t=1997	t0=1999 t=2003	t0=1968 t=2003
Past Exit		17362	18914	21726	22810	29055	30119	19389
Future Entry		11743	9751	7944	5461	3621	0	0
Recent Entry	Sample born in	1506	418	351	295	607	1340	8320
	Sample mover in	0	492	41	26	66	138	576
	Nonsample	2471	1056	615	471	897	2057	6678
	Sample new/ Sample appear	11	26	0	1691	270	2	1534
Reentry/recontact		1183	1759	121	163	450	556	0
Recent Exit	Died between the prior and the current wave	457	155	99	125	159	429	3481
	Responded in the prior wave and nonresponse in the current wave	2854	1396	1148	6570	1377	1887	9565
Response	Responded in the prior and the current wave	17140	20760	22682	17115	18225	18199	5184
Total		54727	54727	54727	54727	54727	54727	54727

Latino sample is excluded from the analysis reported in this table.

Table3 Adjusted and carry over weights assignment rules

Group	Stratum	Data collected at t_0?	Data collected at t?	Adjusted weights Step 1	Adjusted weights Step 2	Carry over weights Step 1	Carry over weights Step 2
Past Exit		No	No	Zero		Zero	
Future Entry		No	No	Zero		Zero	
Recent Entry	Sample born in	No	Yes		Average of head's and wife's weights if there are head and wife in the family and 1/2 of head's weight if there is no wife in the family.		Average of head's and wife's weights if there are head and wife in the family and 1/2 of head's weight if there is no wife in the family.
	Sample mover in	No	Yes		Average of head's and wife's weights if there are head and wife in the family and 1/2 of head's weight if there is no wife in the family.		Average of head's and wife's weights if there are head and wife in the family and 1/2 of head's weight if there is no wife in the family.
	Nonsample	No	Yes	Zero		Zero	
	Sample new/ Sample appearer	No	Yes		a) If a member of a newly added sample then the weight is inverse probability of selection; b) If a sample appear then the weight is equal to an average of sample family members individual weights		a) If a member of a newly added sample then the weight is inverse probability of selection; b) If a sample appear then the weight is equal to an average of sample family members individual weights
Reentry/recontact		No	Yes	Reference weight		Reference weight	
Recent Exit	Died between the prior and the current wave	Yes	No	Zero		Zero	
	Responded in the prior wave and nonresponded in the current wave	Yes	No	Zero		Zero	
Response	Responded in the prior and the current wave	Yes	Yes	weight at t_0*attrition adjustment		*weight at t_0*	

Table 4:1993-2005 Core (Core/Immigrant) Longitudinal Weights construction approach

Sample	1993: t0=1989 t=1993	1994: t0=1993 t=1994	1995: t0=1994 t=1995	1996: t0=1995 t=1996	1997: t0=1996 t=1997	1999: t0=1997 t=1999	2001: t0=1999 t=2001	2003: t0=1999 t=2003	2005: t0=2003 t=2005
Core	Adjusted	Carry over	Carry over	Carry over	Adjusted with the weighting cell process	Carry over			
Immigrant					Weights are inverse prob of selection	Carry over and inverse prob of selection			
Core/Immigrant					Core and Immigrant weights are combined with the 93:7 ratio	Core and Immigrant weights are combined with the 93:7 ratio	Carry over	Adjusted	Carry over

Table 5: Multinomial Logistic Regression, $t_0 = 1989$, $t = 1993$.

(Omitted Category is "Response.")

	Dead			Nonresponse		
Variable	Coefficient		SE	Coefficient		SE
Intercept	-3.2838	***	1.1725	1.6578	***	0.5963
1st income percentile	-1.8266		1.3289	-2.3808	***	0.7073
Log of income	-0.3750	***	0.1030	-0.4189	***	0.0566
100th income percentile	0.7826		0.5762	0.0984		0.3767
Age	0.0380	*	0.0201	-0.0017		0.0072
Age squared	0.0004	**	0.0002	-0.0001		0.0001
North Central	-0.2218		0.2305	0.0285		0.1433
South	-0.1251		0.2215	0.2100		0.1380
West	-0.1912		0.2634	0.2867	*	0.1555
Male	0.7362	***	0.1650	0.1116		0.0930
SEO sample	-1.1320		1.5809	-1.4120	**	0.7112
SEO sample*1st income percentile	2.0049		1.5514	1.4778	*	0.8140
SEO sample*log income	0.1675		0.1406	0.2547	***	0.0675
SEO*Age	-0.0002		0.0279	0.0119		0.0096
SEO*Age squared	0.0000		0.0003	-0.0003	*	0.0001
SEO*North Central	0.4353		0.4333	-0.6119	***	0.1924
SEO*South	0.2722		0.3847	-0.9105	***	0.1737
SEO*West	-0.1842		0.4978	-0.7417	***	0.2070
SEO*Male	-0.2508		0.2406	0.0925		0.1172
Summary statistics:						
Number of observations	15558					
Response profile:						
Dead	359					
Nonresponse	1437					
Response	13762					
Likelihood Ratio:						
Chi-squared	1377.2					
DF	36					

Table 6: Multinomial Logistic Regression, $t_0 = 1999$, $t = 2003$

(Omitted Category is "Response.")

	Dead			Nonresponse		
Variable	Coefficient		SE	Coefficient		SE
Intercept	-2.5325	**	1.0074	-3.0883	***	0.6031
1st income percentile	-2.4680	***	0.8082	0.8540		0.5915
log of income	-0.5134	***	0.0806	-0.0065		0.0542
100th income percentile	1.0293		0.6535	0.0961		0.3850
Age	0.0602	***	0.0195	0.0072		0.0072
Age squared	0.0002		0.0002	-0.0002	**	0.0001
North Central	0.1214		0.2056	0.1581		0.1349
South	0.2181		0.1977	0.1249		0.1295
West	-0.2322		0.2411	-0.2199		0.1359
Male	0.6741	***	0.1414	0.1660	*	0.0853
SEO sample	-3.5952	*	2.0051	1.1743		0.9970
SEO sample*1st income percentile	-7.6510		163.5000	-0.9156		0.8434
SEO sample*log income	0.4892	***	0.1798	-0.1973	**	0.0888
SEO*Age	-0.0509		0.0317	0.0296	*	0.0162
SEO*Age squared	0.0005	*	0.0003	-0.0006	**	0.0003
SEO*North Central	-0.0794		0.6149	0.5619		0.4106
SEO*South	-0.2502		0.5553	0.5027		0.3910
SEO*West	-0.9581		0.9320	0.7212		0.4728
SEO*Male	-0.0136		0.3066	0.2908	*	0.1602
Immigrant sample	-1.0665	***	0.3695	1.2554	***	0.1015
Summary statistics						
Number of observations	15292					
Response profile:						
Dead	366					
Nonresponse	842					
Response	14084					
Likelihood Ratio:						
Chi-squared	1581.6					
DF	38					

Table 7: the Core (Core/Immigrant)
Longitudinal Individual Weights
count

		Core sample (SRC, SEO) and Immigrant sample						Latino sample					
Year	Total number of individuals in the study	Total number of individuals	Total number of sample individuals	Total number of nonsample individuals	Number of cases with positive individual weight	Number of cases with zero individual weight	Number of cases with missing individual weight	Total number of individuals	Total number of sample individuals	Total number of non-sample individuals	Number of cases with positive individual weight	Number of cases with zero individual weight	Number of cases with missing individual weight
1989	20451	20451	15559	4892	15563	4886	2	0	0	0	0	0	0
1990	28197	20745	15621	5124	15622	5121	2	7452	7116	336	0	7452	0
1991	27845	20770	15602	5168	15602	5166	2	7075	6750	325	0	7075	0
1992	29275	21145	15749	5396	15749	5393	3	8130	7396	734	0	8130	0
1993	29726	22311	16118	6193	16118	6193	0	7415	6500	915	0	7415	0
1994	31546	24512	18158	6354	18158	6354	0	7034	6182	852	0	7034	0
1995	29884	23929	17705	6224	17705	6224	0	5955	5273	682	0	5955	0
1996	23810	23810	17593	6217	17593	6217	0	0	0	0	0	0	0
1997	19761	19761	15051	4710	15051	4710	0	0	0	0	0	0	0
1999	20515	20515	15317	5198	15317	5198	0	0	0	0	0	0	0
2001	21400	21400	15646	5754	15646	5754	0	0	0	0	0	0	0
2003	22290	22290	16012	6278	16012	6278	0	0	0	0	0	0	0
2005	22918	22918	16620	6298	16620	6298	0	0	0	0	0	0	0

Table 8: the Core
(Core/Immigrant) Longitudinal
Family Weights count

		Core sample (SRC, SEO) and Immigrant sample					Latino sample				
YEAR	Total number of families	Total number of families	Number of families with no sample person	Number of families with positive weight	Number of families with zero weight	Number of families with missing weight	Total number of families	Number of families with no sample person	Number of families with positive weight	Number of families with zero weight	Number of families with missing weight
1989	7114	7114	1	7114	0	0	0	0	0	0	0
1990	9371	7328	82	7249	79	0	2043	0	0	2043	0
1991	9363	7375	95	7283	92	0	1988	0	0	1988	0
1992	9829	7561	100	7465	96	0	2268	1	0	2268	0
1993	9977	7873	92	7781	92	0	2104	0	0	2104	0
1994	10765	8658	195	8463	195	0	2107	1	0	2107	0
1995	10401	8567	201	8366	201	0	1834	4	0	1834	0
1996	8511	8511	131	8380	131	0	0	0	0	0	0
1997	6747	6747	121	6626	121	0	0	0	0	0	0
1999	6997	6997	146	6851	146	0	0	0	0	0	0
2001	7406	7406	211	7195	211	0	0	0	0	0	0
2003	7822	7822	257	7565	257	0	0	0	0	0	0
2005	8002	8002	0	8002	0	0	0	0	0	0	0

Table 9: Distribution of the Core (Core/Immigrant) Longitudinal Individual Weights (sample individuals)

Year	Number of observations	Mean	Standard deviation	Minimum	Maximum	Coefficient of variation
1989	15564	21.06	16.61	0.4	99.9	0.79
1990	15623	20.96	16.56	0.43	131.88	0.79
1991	15604	20.81	16.42	0.28	131.88	0.79
1992	15751	20.71	16.42	0.28	131.88	0.79
1993	16118	21.76	17.11	0.28	109.92	0.79
1994	18158	20.51	16.78	0.28	109.92	0.82
1995	17705	20.42	16.7	0.28	109.92	0.82
1996	17593	20.23	16.58	0.28	109.92	0.82
1997	15051	26.35	19.51	0.29	167.68	0.74
1999	15317	25.67	19.05	0.29	167.68	0.74
2001	15646	25.07	18.97	0.25	167.68	0.76
2003	16012	25.62	19.54	0.25	173.56	0.76
2005	16620	24.81	19.33	0.23	173.56	0.78

Table 10: Distribution of the Core (Core/Immigrant) Longitudinal Family Weights (families with no sample person are excluded)

Year	Number of observations	Mean	Standard deviation	Minimum	Maximum	Coefficient of variation
1989	7114	19.41	15.54	0.4	99.9	0.80
1990	7249	19.38	15.51	0.41	110.53	0.80
1991	7283	19.26	15.44	0.28	110.53	0.80
1992	7465	19.15	15.47	0.25	115.39	0.81
1993	7781	18.5	15.54	0.10	107.83	0.84
1994	8463	17.95	15.21	0.10	107.83	0.85
1995	8366	17.78	15.08	0.10	107.83	0.85
1996	8380	17.52	14.94	0.10	107.83	0.85
1997	6626	24.13	17.40	0.11	167.68	0.72
1999	6851	23.04	16.97	0.06	167.68	0.74
2001	7195	22.03	16.74	0.06	167.68	0.76
2003	7565	22.06	17.06	0.12	132.64	0.77
2005	8002	21.04	16.82	0.12	136.03	0.80

Table 11: Comparison of PSID and CPS Estimates of Age

A. Household head

	PSID unweighted		PSID weighted		CPS weighted		Ratio	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Year	[1]	[2]	[3]	[4]	[5]	[6]	[3]/[5]	[4]/[6]
1990	44.1	39	47.8	43	47.9	45	1.00	0.96
1991	44.5	40	47.9	44	48.1	45	1.00	0.98
1992	44.7	40	47.9	44	48.3	45	0.99	0.98
1993	44.5	40	48.4	45	48.3	45	1.00	1.00
1994	44.6	41	48.4	45	48.2	45	1.01	1.00
1995	44.7	41	48.6	45	48.3	45	1.01	1.00
1996	44.5	41	48.8	46	48.4	45	1.01	1.02
1997	44.4	41	48.3	45	48.4	46	1.00	0.98
1999	44.4	42	48.8	46	48.5	46	1.00	1.00
2001	44.9	43	49.3	47	48.7	46	1.01	1.02
2003	44.9	43	49.6	48	48.6	47	1.02	1.02
2005	45.1	44	49.9	48	49.0	47	1.02	1.02

B. All individuals

	PSID unweighted		PSID weighted		CPS weighted		Ratio	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Year	[1]	[2]	[3]	[4]	[5]	[6]	[3]/[5]	[4]/[6]
1990	30.0	28	35.1	33	34.4	32	1.02	1.03
1991	30.2	29	35.3	33	34.5	33	1.02	1.00
1992	30.4	29	35.5	34	34.6	33	1.03	1.03
1993	30.3	29	35.5	34	34.7	33	1.02	1.03
1994	30.4	29	35.2	34	34.4	33.	1.02	1.03
1995	30.6	29	35.4	34	34.6	33	1.02	1.03
1996	30.5	29	35.7	35	34.6	33	1.03	1.06
1997	29.8	28	35.0	34	34.8	34	1.00	1.00
1999	30.4	28	35.7	35	35.1	34	1.02	1.03
2001	30.8	29	36.3	36	35.6	35	1.02	1.03
2003	31.2	29	36.5	36	35.8	35	1.02	1.03
2005	31.4	29	36.9	36	36.1	36	1.02	1.00

Table 12: Comparison of PSID and CPS Estimates of Gender

	PSID unweighted		PSID weighted		CPS weighted		Ratio	
Year	Male	Female	Male	Female	Male	Female	Male	Female
	[1]	[2]	[3]	[4]	[5]	[6]	[3]/[5]	[4]/[6]
	%	%	%	%	%	%		
1992	48.0	51.9	47.5	52.4	48.7	51.2	0.98	1.02
1993	47.8	52.2	47.7	52.2	48.7	51.2	0.98	1.02
1994	48.1	51.8	48.0	51.9	48.8	51.1	0.98	1.02
1995	47.9	52.0	47.9	52.0	48.8	51.1	0.98	1.02
1996	47.9	52.0	47.8	52.1	48.8	51.1	0.98	1.02
1997	47.9	52.0	48.0	51.9	48.9	51.0	0.98	1.02
1999	48.0	51.9	48.1	51.8	48.8	51.1	0.99	1.01
2001	47.9	52.0	48.0	51.9	48.8	51.1	0.98	1.02
2003	47.9	52.0	48.1	51.8	48.9	51.0	0.98	1.01
2005	47.8	52.1	48.2	51.7	49.0	50.9	0.98	1.02

Table 13: Comparison of PSID and CPS Estimates of Race

A. Household data: Race of head

	PSID unweighted		PSID weighted		CPS weighted		Ratio	
Year	Non-black	Black	Non-black	Black	Non-black	Black	Non-black	Black
	[1]	[2]	[3]	[4]	[5]	[6]	[3]/[5]	[4]/[6]
	%	%	%	%	%	%		
1990	63.3	36.7	86.2	13.8	88.8	11.2	0.97	1.23
1991	63.3	36.7	86.0	14.0	88.7	11.3	0.97	1.24
1992	63.4	36.6	86.2	13.8	88.4	11.6	0.98	1.19
1993	63.8	36.2	86.8	13.2	88.4	11.6	0.98	1.14
1994	64.5	35.5	86.4	13.6	88.4	11.6	0.98	1.17
1995	64.5	35.5	86.5	13.5	88.2	11.8	0.98	1.14
1996	64.6	35.4	86.6	13.4	88.4	11.6	0.98	1.16
1997	69.8	30.2	87.4	12.6	88.0	12.0	0.99	1.05
1999	69.8	30.2	87.4	12.6	87.9	12.1	0.99	1.04
2001	69.6	30.4	87.4	12.6	87.8	12.2	1.00	1.03
2003	68.4	31.6	87.2	12.8	87.9	12.1	0.99	1.06
2005	66.7	33.3	86.1	13.9	87.8	12.2	0.98	1.14

B. Individual data: individual race is approximated by the race of head in the case of PSID data

	PSID unweighted		PSID weighted		CPS weighted		Ratio	
year	Non-black	Black	Non-black	Black	Non-black	Black	Non-black	Black
	[1]	[2]	[3]	[4]	[5]	[6]	[3]/[5]	[4]/[6]
	%	%	%	%	%	%		
1990	61.4	38.6	86.0	14.0	87.7	12.3	0.98	1.14
1991	61.4	38.6	85.9	14.1	87.6	12.4	0.98	1.14
1992	61.9	38.1	86.1	13.9	87.5	12.5	0.98	1.11
1993	62.1	37.9	86.1	13.9	87.4	12.6	0.99	1.10
1994	62.8	37.2	85.8	14.2	87.3	12.7	0.98	1.12
1995	63.2	36.8	86.1	13.9	87.2	12.8	0.99	1.09
1996	63.4	36.6	86.0	14.0	87.2	12.8	0.99	1.09
1997	66.8	33.2	86.7	13.3	87.2	12.8	0.99	1.04
1999	66.8	33.2	86.9	13.1	87.1	12.9	1.00	1.02
2001	67.0	33.0	86.9	13.1	87.3	12.7	1.00	1.03
2003	66.1	33.9	86.6	13.4	87.5	12.5	0.99	1.07
2005	64.6	35.4	86.0	14.0	87.4	12.6	0.98	1.11

Table 14: Comparison of PSID and CPS estimates of Family income

	PSID unweighted		PSID weighted		CPS weighted		Ratio	
Year	Median	Mean	Median	Mean	Median	Mean	Median	Mean
	[1]	[2]	[3]	[4]	[5]	[6]	[3]/[5]	[4]/[6]
	$	$	$	$	$	$		
1990	41,915	53,994	45,429	59,323	45,555	56,220	1.00	1.06
1991	40,719	52,803	43,109	56,818	44,678	54,741	0.96	1.04
1992	40,178	52,179	43,018	56,592	43,018	53,445	1.00	1.06
1993	40,413	53,199	42,888	57,322	42,457	53,120	1.01	1.08
1994	39,240	53,936	41,942	59,144	41,931	53,164	1.00	1.11
1995	40,193	54,712	42,934	59,884	42,414	53,857	1.01	1.11
1996	41,146	55,139	43,672	59,491	43,571	57,589	1.00	1.03
1997	43,619	58,428	44,500	61,305	43,805	58,666	1.02	1.04
1999	45,554	62,245	47,231	65,421	46,552	62,126	1.01	1.05
2001	47,924	67,059	48,945	70,944	47,634	64,805	1.03	1.09
2003	44,746	60,926	45,852	64,372	46,064	62,795	1.00	1.03
2005	46,204	64,282	46,711	68,611	45,712	62,581	1.02	1.10

Table 15: Names of the Core and Core/Immigrant Longitudinal Weight Variables, 1993-2005

	Core Longitudinal weight		Core/Immigrant Longitudinal weight	
Year	Individual	Family	Individual	Family
1993	ER30864	V23361		
1994	ER33119	ER4160		
1995	ER33275	ER7000		
1996	ER33318	ER9251		
1997			ER33430	ER12084
1999			ER33546	ER16518
2001			ER33637	ER20394
2003			ER33740	ER24179
2005			ER33848	ER28078