**Technical Report**

Panel Study of Income Dynamics

Construction and Evaluation of the Longitudinal Sample Weight 2007

**August, 2009**

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

This document describes the construction of the 2007 Core/Immigrant individual and family longitudinal sample weights for the Panel Study of Income Dynamics (PSID) and compares its distribution on key estimates to those obtained from the Current Population Survey.

## 1. The PSID Sample and Following Strategy in 2007

PSID interviewed 8,289 families in 2007. All of these families are members of the Core or the Immigrant samples[1]. In 2007, the following rule was the same as in the prior 2005 wave[2]. Specifically, sample persons who participated in the previous wave survey were followed. Additionally, the PSID attempted to obtain an interview with sample individuals who did not respond in the prior wave (2005 survey year), but responded in the wave before the prior wave (2003 survey year).

---

[1] The Core sample is the combination of two samples: 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 respondents from the Core sample have been interviewed since 1968. In 1997 and 1999 a sample of the post-1968 immigrants was added. Information on the immigrant sample is available under the column "Other Documentation" for years 1997 and 1999 at the following url: http://simba.isr.umich.edu/Zips/ZipMain.aspx.

[2] For more detail on the following rules in 1993-2005 survey years see Table 1 in Gouskova et al. (2008) (http://psidonline.isr.umich.edu/data/weights/Long-weights-doc.pdf)

## 2. Methodological Approach to the 2007 Core/Immigrant Longitudinal Weight Construction

The 2007 weights are attrition-adjusted weights. The last attrition adjustment of the PSID longitudinal weights was done in 2003. Thus, the construction of the 2007 longitudinal weights starts with the 2003 longitudinal weight as the basis. The 2007 weights are constructed using the same methodological approach as used for the 2003 longitudinal weights, i.e. for those who responded in 2003 the 2007 weights are obtained by multiplying the 2003 longitudinal weight by the attrition adjustment factor, and for new sample members, (sample newborns and those sample persons who moved in) the 2007 weight is calculated as an average of head's and wife's weight in 2003. The detailed description of this approach is provided in Gouskova et al. (2008). To account for attrition between the 2003 and 2007 waves among the sample respondents who participated in the 2003 study, an adjustment factor was calculated. Table 1 reports the results of the multinomial logistic regression estimating probabilities of three possible states in 2007 for those who responded in 2003: response, non- response and known death. Using the estimates, the probability of response conditional on surviving has been calculated as described in section 4.3 in Gouskova et al. (2008). The inverse of this probability is the adjustment factor that is used to multiply the 2003 individual weight in order to obtain the 2007 individual weight.

## 3. Results

### 3.1 Descriptive Statistics of the Core/ Immigrant Longitudinal Weights, 2007

Tables 2 through Table 5 describe the resulting 2007 Core/ Immigrant longitudinal weight. To enable comparison of the longitudinal weights across years, the same set of descriptors is reported for the longitudinal weights from the three prior waves.

Tables 2 and 3 compare the total number of cases with positive, zero, and missing values for individual and family weights with the total numbers of sample and non-sample individuals (families with and without sample members). For individual weights, the number of weights with a positive value is equal to the number of sample persons, and the number of the zero-valued individual weights is the same as the number of non-

sample persons (Table 2). As with the 2005 survey, in 2007 all families had at least one sample member (Table 3). As a result all PSID families in 2007 carry a positive longitudinal family weight.

Tables 4 and 5 report summary statistics for the longitudinal individual and family weights. The summary statistics of the 2007 longitudinal weights are close to those in the three most recent surveys. Across years, the measures of dispersion indicate that there is an increasing trend in variability of the distribution in the individual and family weights.

*3.2 Distributions of Some Demographic and Socioeconomic Characteristics, PSID and CPS*

Tables 6 through Table 9 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) for the 2001 – 2007 PSID survey years. These tables are useful for examining three features of the PSID data: consistency of unweighted and weighted estimates across years, the effect of the longitudinal weights on the distributions of the characteristics, and, finally, the closeness of the PSID estimates with those obtained from the CPS data[3].

The tables show that consistency across years of the weighted distributions is comparable to the consistency of the unweighted distributions. Comparison of the unweighted and weighted PSID distributions with the CPS distributions reveals that in a majority of cases the weighted estimates are closer to CPS estimates than are the estimates obtained without weights. In addition, the weighted PSID and CPS estimates align fairly closely in most cases.

The 2007 longitudinal weights are stored in the PSID data archive under names provided in Table 10.

---

[3] 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, 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.

References

Gouskova E., S. Heeringa, K. McGonagle, R. Schoeni, and F. Stafford, 2008. "Panel Study of Income Dynamics Revised Longitudinal Weights 2007", PSID Technical Report #08-05, ISR, University of Michigan. http://psidonline.isr.umich.edu/Publications/Papers/tsp/2008-05_PSID_Revised_Longitudinal_Weights_1993-2005%20.pdf

Table 1: Multinomial Logistic Regression, $t_0 = 2003$, $t = 2007$ (Omitted Category is "Response.")

| | 2007 status: dead | | 2007 status: non-response | | |
|---|---|---|---|---|---|
| | estimate | | std | estimate | | std |
| Intercept | -2.9759 | *** | 0.9777 | -3.4647 | *** | 0.5781 |
| Income 1st percentile | -1.4974 | ** | 0.7422 | 0.6198 | | 0.6236 |
| Log of income | -0.3391 | *** | 0.0823 | 0.0743 | | 0.0512 |
| Income 100th percentile | -10.7137 | | 208.4000 | -0.4711 | | 0.3813 |
| Age | 0.0216 | | 0.0165 | 0.0086 | | 0.0067 |
| Age squared | 0.0005 | *** | 0.0001 | -0.0002 | ** | 0.0001 |
| North Central (0/1) | 0.0128 | | 0.2083 | -0.6685 | *** | 0.1304 |
| South (0/1) | -0.3616 | * | 0.2131 | -0.1527 | | 0.1122 |
| West (0/1) | -0.0382 | | 0.2282 | -0.4785 | *** | 0.1206 |
| D male (0/1) | 0.3809 | ** | 0.1485 | 0.0751 | | 0.0812 |
| D SMSA (0/1) | -0.0819 | | 0.1507 | -0.0127 | | 0.0886 |
| D might move (0/1) | 0.1124 | | 0.1872 | -0.0146 | | 0.0897 |
| SEO (0/1) | -2.0109 | | 1.8157 | -1.1819 | | 1.0145 |
| SEO*(income 1st percentile ) | 1.0831 | | 1.1108 | 0.4548 | | 0.8938 |
| SEO*(Log of income) | 0.0634 | | 0.1490 | 0.0014 | | 0.0898 |
| SEO*(Age) | 0.0678 | * | 0.0346 | 0.0194 | | 0.0149 |
| SEO*(Age squared) | -0.0005 | * | 0.0003 | -0.0003 | | 0.0002 |
| SEO*(North Central(0/1)) | -0.1003 | | 0.5532 | 0.6148 | * | 0.3458 |
| SEO*(South (0/1)) | 0.0135 | | 0.5145 | 0.3109 | | 0.3093 |
| SEO*(West (0/1)) | -0.3951 | | 0.7228 | 0.8487 | ** | 0.3711 |
| SEO*(D male (0/1)) | 0.2335 | | 0.2921 | 0.3176 | ** | 0.1570 |
| SEO*(D SMSA (0/1)) | -0.5034 | | 0.3186 | 0.1763 | | 0.1780 |
| SEO*(D might move (0/1)) | 0.4507 | | 0.3360 | 0.4450 | *** | 0.1626 |
| Immigrant sample (0/1) | -0.5890 | * | 0.3183 | 1.0140 | *** | 0.0998 |

| *Summary statistics* | | | | | | |
|---|---|---|---|---|---|---|
| N: | 16005 | | | | | |

| *Response profile:* | | | | | | |
|---|---|---|---|---|---|---|
| Dead | 330 | | | | | |
| Non-response | 902 | | | | | |
| Response | 14773 | | | | | |

| *Likelihood Ratio:* | | | | | | |
|---|---|---|---|---|---|---|
| Chi-squared | 1313.4 | | | | | |
| DF | 46 | | | | | |

Table 2: The Core/Immigrant
Longitudinal Individual Weights,
2001- 2007

| | | Core sample (SRC, SEO) and Immigrant sample | | | | |
|---|---|---|---|---|---|---|
| Year | Total number of individuals in the study | 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 |
| 2001 | 21400 | 15646 | 5754 | 15646 | 5754 | 0 |
| 2003 | 22290 | 16012 | 6278 | 16012 | 6278 | 0 |
| 2005 | 22918 | 16620 | 6298 | 16620 | 6298 | 0 |
| 2007 | 23508 | 16906 | 6602 | 16906 | 6602 | 0 |

Table 3: The Core/Immigrant  Longitudinal
Family Weights, 2001- 2007

| YEAR | Total number of families | Core sample (SRC, SEO) | | | |
|------|---------|---------|---------|---------|---------|
| | | 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 |
| 2001 | 7406 | 211 | 7195 | 211 | 0 |
| 2003 | 7822 | 257 | 7565 | 257 | 0 |
| 2005 | 8002 | 0 | 8002 | 0 | 0 |
| 2007 | 8289 | 0 | 8289 | 0 | 0 |

Table 4: Summary statistics of the Core/Immigrant Longitudinal Individual
Weights, 2001- 2007
(sample individuals only)

| Year | Number of observations | Mean | Standard deviation | Minimum | Maximum | Coefficient of variation |
|------|------------------------|-------|--------------------|---------|---------|--------------------------|
| 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 |
| 2007 | 16906 | 25.38 | 20.09 | 0.2 | 181.45 | 0.79 |

Table 5: Summary Statistics of the Core/Immigrant Longitudinal Family Weights
Distribution, 2001- 2007
(the 2001 and 2003 results are based on families with positive weights only)

| Year | Number of observations | Mean | Standard deviation | Minimum | Maximum | Coefficient of variation |
|------|------------------------|------|--------------------|---------|---------|--------------------------|
| 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 |
| 2007 | 8289 | 21.32 | 17.4 | 0.1 | 139.34 | 0.82 |

Table 6: Comparison of PSID
and CPS Estimates of Age, 2001-
2007

A. Household head

| | PSID unweighted | | PSID weighted | | CPS weighted | | Ratio of weighted distributions | |
| | Mean | Median | Mean | Median | Mean | Median | Mean | Median |
| Year | [1] | [2] | [3] | [4] | [5] | [6] | [3]/[5] | [4]/[6] |
| 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 |
| 2007 | 45.0 | 44 | 50.1 | 49 | 49.3 | 48 | 1.02 | 1.02 |

B. All individuals

| | PSID unweighted | | PSID weighted | | CPS weighted | | Ratio of weighted distributions | |
| | Mean | Median | Mean | Median | Mean | Median | Mean | Median |
| Year | [1] | [2] | [3] | [4] | [5] | [6] | [3]/[5] | [4]/[6] |
| 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 |
| 2007 | 31.6 | 29 | 37.4 | 37 | 36.4 | 36 | 1.02 | 1.03 |

Table 7: Comparison of PSID and CPS Estimates of Gender, 2001-2007

| Year | PSID unweighted | | PSID weighted | | CPS weighted | | Ratio of weighted distributions | |
|------|------|--------|------|--------|------|--------|------|--------|
| | Male | Female | Male | Female | Male | Female | Male | Female |
| | [1] | [2] | [3] | [4] | [5] | [6] | [3]/[5] | [4]/[6] |
| | % | % | % | % | % | % | | |
| 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 |
| 2007 | 47.9 | 52.1 | 48.6 | 51.4 | 49.1 | 50.9 | 0.99 | 1.01 |

Table 8: Comparison of PSID and CPS Estimates of Race, 2001-2007

A. Household data: Race of head

| Year | PSID unweighted | | PSID weighted | | CPS weighted | | Ratio of weighted distributions | |
|---|---|---|---|---|---|---|---|---|
| | Non-black | Black | Non-black | Black | Non-black | Black | Non-black | Black |
| | [1] | [2] | [3] | [4] | [5] | [6] | [3]/[5] | [4]/[6] |
| | % | % | % | % | % | % | | |
| 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 |
| 2007 | 65.7 | 34.3 | 85.9 | 14.1 | 87.6 | 12.4 | 0.98 | 1.14 |

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

| year | PSID unweighted | | PSID weighted | | CPS weighted | | Ratio of weighted distributions | |
|---|---|---|---|---|---|---|---|---|
| | Non-black | Black | Non-black | Black | Non-black | Black | Non-black | Black |
| | [1] | [2] | [3] | [4] | [5] | [6] | [3]/[5] | [4]/[6] |
| | % | % | % | % | % | % | | |
| 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 |
| 2007 | 64.2 | 35.8 | 85.9 | 14.1 | 87.4 | 12.6 | 0.98 | 1.12 |

Table 9: Comparison of PSID and CPS estimates of Family Income, 2001-2007

| | PSID unweighted | | PSID weighted | | CPS weighted | | Ratio of weighted distributions | |
|---|---|---|---|---|---|---|---|---|
| Year | Median | Mean | Median | Mean | Median | Mean | Median | Mean |
| | [1] | [2] | [3] | [4] | [5] | [6] | [3]/[5] | [4]/[6] |
| | $ | $ | $ | $ | $ | $ | | |
| 2001 | 50,869 | 71,179 | 51,952 | 75,303 | 50,561 | 68,787 | 1.03 | 1.09 |
| 2003 | 47,495 | 64,670 | 48,669 | 68,327 | 48,894 | 66,653 | 1.00 | 1.03 |
| 2005 | 49,043 | 68,232 | 49,581 | 72,827 | 48,521 | 66,426 | 1.02 | 1.10 |
| 2007 | 48,665 | 67,107 | 50,385 | 71,856 | 49,413 | 68,457 | 1.02 | 1.05 |

Notes: 1) Data correspond to a survey year and represent income in the previous calendar year.
2) Nominal values are expressed in 2007 dollars.

Table 10: Names of the Core  and Core/Immigrant
Longitudinal Weight Variables, 1993-2007

| | 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 |
| 2007 | | | ER33950 | ER41069 |