Technical Report

Panel Study of Income Dynamics
2023 PSID Cross-Sectional Individual Weights

June 2026

Wen Chang, Raphael Nishimura, Noura Insolera, and Thomas Crossley

Survey Research Center, Institute for Social Research
University of Michigan,
Ann Arbor, MI

This technical report documents the methodological approach to the cross-sectional weights constructed for the individuals from the 2023 Panel Study of Income Dynamics (PSID-2023).

A separate set of PSID longitudinal analysis weights for individuals and family units are also available, and are documented in Chang et al. (2026). While researchers have always been able to perform cross-sectional analysis using longitudinal individual weights for PSID sample persons, the cross-sectional individual weights offer an additional approach for weighted cross-sectional estimation based on the PSID individual data. Specifically, the PSID cross-sectional individual weights permit analysts to use all available data for both PSID sample and nonsample[1] persons to estimate population characteristics or model population relationships at specific points in time.

In addition, the PSID cross-sectional weights are calibrated to the population characteristics from the Current Population Survey (CPS) or American Community Survey (ACS) for the respective year.[2] CPS and ACS are large, high-quality national survey samples conducted by the U.S. Census Bureau that provide annual population estimates by demographic characteristics in non-census years. This calibration adjustment not only aligns the PSID sample weighted distribution to the population by selected socio-demographic dimensions, but it also has the potential to mitigate bias from nonresponse and coverage error, and improve the precision of survey estimates. PSID has provided the cross-sectional individual weight since 1997 and plans to provide the cross-sectional individual weight for each future wave.

Analyses requiring repeated family-level cross-sectional estimates (e.g., for 2011, 2013, 2015, 2017, 2019, 2021, and 2023) can use the longitudinal family weight for each wave.[3]

This technical report is organized into four sections. Section I defines sample and nonsample persons in PSID and explains the rationale for creating the cross-sectional weights. The "fair shares" methodology that underlies the construction of the PSID cross-sectional weights is discussed in Section II. Section III describes how the

1 The PSID nonsample persons have longitudinal individual weights equal to zero so they would not be included in the weighted analysis using longitudinal individual weights.

2 The PSID longitudinal weights are not calibrated at each wave against external, nationally representative population estimates. The population characteristics used for calibration were based on CPS estimates from 1997 to 2013 and have been based on ACS estimates since 2015.

3 The 2017 new immigrant sample have zero longitudinal family weight in 2017 so they would be excluded from the estimates in 2017.

cross-sectional weights are constructed and provides a descriptive analysis of the weights. Recommendations for using the weights are provided in Section IV.

I. Background

Sample and nonsample persons in the PSID

PSID traditionally categorizes individuals into one of two groups: sample persons and nonsample persons. The definition of these categories has changed slightly over the years. From 1968 to 1993, a sample person was defined as someone who was either an original sample person, i.e., resident of a PSID sample family in 1968, or offspring born to or adopted by a sample individual who was actively participating in the study at the time. A newborn child had to appear in the study at the wave immediately following their birth to be considered a sample person. In 1994, the definition of a sample person was expanded to also include children born to or adopted by a sample person even when the sample person was not participating in the study, i.e., the child need not have been residing with a responding panel family unit at birth or adoption.

In 1997, a baseline sample of post-1968 immigrant families and individuals was added to PSID. The same current PSID definition of sample persons (implemented in 1994) applies to the 1997 New Immigrant Immigrant Supplement (NIS-1997). For NIS-1997 families, the true baseline year for sample selection and sample status determination is 1997 or 1999.

In 2017, a baseline sample of post-1997 immigrant families and individuals was added to PSID. The same current PSID definition of sample persons also applies to the 2017 Immigrant sample. The 2017 Immigrant sample was recruited by either the New Immigrant Supplement (NIS-2017) study or the New Immigrant Multiplicity Supplement (NIMS) study. The NIMS study was needed because of the concurrent recruitment with the Health and Retirement Study (HRS). See Chang et al. (2021) for the details of the sample recruitment of the 2017 Immigrant sample. For 2017 Immigrant families responding in 2017, the sample status determination for individuals is 2017. For NIMS families and NIS-2017 families that did not respond in 2017 but responded in 2019, the sample status determination year is 2019.

All other members of PSID families are considered nonsample persons. They are typically spouses or partners and other family unit members. See Johnson et al. (2018) for a detailed background on the PSID. Under the conventional method for computing PSID longitudinal weights for individuals, nonsample persons are

automatically assigned a “0” weight and, thus, excluded from any properly weighted longitudinal analysis of the PSID individual data. The justification for assigning a zero longitudinal weight to nonsample persons was two-fold. First, barring any biases due to nonresponse and attrition, the dynamic sampling design for individuals and families employed in PSID provides unbiased representation of the survey population at each measurement point (cross-sectionally) and over time (longitudinally). Second, the process of dynamic recruitment of nonsample persons to PSID families is left-censored. This means that the time at which a nonsample person is first observed in a longitudinal sequence of observations is stochastic—it is potentially dependent on age and other factors, but is otherwise random after conditioning on such covariates.

Rationale for creating the cross-sectional weights

The data loss resulting from excluding nonsample persons was not significant in the early years because nonsample persons represented a modest fraction of the total individuals in the PSID sample of families. For instance, among 17,212 total PSID individuals in 1969, 537 (3.1%) were nonsample persons. However, as appendix Table A1 shows, with the passage of time, nonsample persons have comprised an increasing and now substantial share of the total PSID individuals. By 2023, the number of nonsample persons grew to 6,506 out of 24,536 PSID individuals (26.5%).

Although PSID supports various forms of longitudinal analyses, cross-sectional analysis is a popular usage of the PSID data. In order to increase the effective sample size for such analysis, a new set of weights has been developed at the individual level since 1997. These weights are labeled cross-sectional weights to underscore their purpose and to distinguish them from the traditional PSID longitudinal weights. Unlike the longitudinal individual weights, the cross-sectional individual weights are non-zero for both sample and nonsample persons. This allows information on sample and nonsample persons to be included in weighted analyses.

For the waves prior to 1997, data users are advised to use longitudinal weights to conduct cross-sectional analyses, recognizing that for these earlier years the analysis will be based only on PSID sample persons.

II. “Fair Shares” Methodology for Constructing PSID Cross-sectional Weights

As early as 1984, statisticians working in the U.S. Survey of Income and Program Participation (SIPP) began to study weighting methodologies for including “nonsample” persons who entered a dynamic, longitudinal sample (Huang, 1984). In 1987, the PSID Board of Overseers expressed interest in a methodology for incorporating the

increasing number of nonsample persons in PSID families into weighted cross-sectional analyses that would represent the general population. Kalton (1987) and Little (1989) developed working papers for the PSID Board that looked specifically at methodology that would enable both PSID sample and nonsample persons to be included in cross-sectional analyses of the panel data. Subsequently, several major panel studies modeled on the PSID and its "dynamic sampling" method have employed the methods discussed in these early papers to develop a cross-sectional weight for point-in-time analyses of the panel data. These include the British Household Panel Survey (Lynn, et al., 2006) and the Canadian Survey of Labour and Income Dynamics (Lavallee, 1995). A comprehensive review of the theory and methods for cross-sectional weight development in longitudinal surveys is provided by Kalton and Brick (1995) and Ernst (1989).

Following Kalton and Brick (1995), one method for assigning nonzero weights to all members—both sample and nonsample persons—of a PSID family unit is labeled the "fair shares" method. Application of the fair shares method assumes that the probability of observing each person in a family unit is equal to the probability of observing the family unit itself. This equivalence of family unit and individual probabilities was true for the original samples of PSID family units and individuals first interviewed in the 1968 baseline wave. However, in subsequent waves, probabilities for nonsample persons who were not members of a 1968 sample family unit were unknown or could not be readily determined.

At any data collection time point, *t*, an initial non-zero cross-sectional weight for each person (W_{it}^{0})

in a PSID family unit can be assigned using the fair shares method:

$$W_{i,t}^{0} = \sum_{i=1}^{n_f} \alpha_i \cdot w_{i,t}^{*}$$

Where:

n_f = total number of sample and nonsample persons in family f;

$w_{i,t}^{*}$ = the current non-zero individual weight for sample person, I, at wave t

=0 if person i is nonsample;

α_i = (general) an arbitrary influence weight $\in$ (0,1), $\sum_{i=1}^{n_f} \alpha_i = 1$.

In general, the values of α_i may be derived to optimize the precision of a specific population estimator (e.g., a population total); however, here we choose an equal individual weighting scheme, with $\alpha_i = 1/n_f$. This choice means that $W_{i,t}^0$ for all family members is equal to the PSID Longitudinal Family Unit Weight at wave *t*.

Using a version of the "fair shares" methodology described above, cross-sectional weights for all PSID individuals have been constructed for every wave since 1997.

III. Weight construction

2023 Cross-Sectional Individual Weight

As described in Section II, we chose an equal individual weighting scheme for the Fair Shares method, so the initial non-zero cross-sectional weight for each person is equivalent to the PSID longitudinal family weight at wave *t*. We then calibrate those initial weights to the population totals estimated from the 2023 ACS 1-year Public Use Microdata Sample (PUMS) data.

In order to decrease bias from nonresponse and coverage error while not increasing sampling variance of the survey estimates, the covariates used in the calibration adjustment should be correlated with both the survey response and the study outcomes (Little and Vartivarian, 2005). For this reason, we started to use a different approach to select the calibration variables since 2017 that ensures that calibration variables are ones that are correlated with both survey response and survey outcome variables.

The following key PSID survey outcomes variables that were used to select calibration variables are:

- mean age of reference person
- percentage of reference persons with health insurance
- percentage of spouse/partner with health insurance
- mean family unit income
- mean reference person labor income
- mean family unit wealth
- percentage of black - reference person
- percentage of white - reference person
- percentage of foreign born - reference person
- percentage of foreign born - spouse/partner
- percentage of families owning a home
- percentage of families with food stamps/SNAP

- mean total food spending

To select the variables (and interactions) that were employed in the calibration, the 13 selected outcome variables from the 2023 PSID interview were regressed on demographic and socio-economic characteristics for each individual.

The regression model for each of the 13 outcomes included all main and two-way interaction effects for each of the following predictors:

- age of individual (0-9/10-19/20-29/30-39/40-49/50-59/60-69/70+)
- sex of individual (Male/Female),
- race of reference person (Black alone or in combination with one or more other races /Non-Black)
- race of reference person (Asian alone or in combination with one or more other races /Non-Asian)
- ethnicity of reference person (Hispanic/Non-Hispanic)
- region (Northeast/Midwest/South/West)
- education of individual (24 years old or younger/11 years or less/12 years/13-15 years/16 years/17 years or more)
- family unit type and employment status (FES) (LF=Labor Force)
    - family unit headed by a couple: reference person and spouse/partner in LF
    - family unit headed by a couple: reference person or spouse/partner in LF
    - family unit headed by a couple: Neither reference person nor spouse/partner in LF
    - male reference person, no spouse/partner present, in LF
    - female reference person, no spouse/partner present, in LF
    - Non-couple, reference person not in LF
- family unit size (1/2/3/4 or more)
- presence of children (Yes/No)
- family unit with foreign-born reference person or foreign-born spouse/partner (Yes/No)

Using the results of these 13 regression models, the final set of controls for the weight calibration was chosen to include all the main effects (regardless of their level of explanation on the survey outcomes) and any two-way interactions of these predictors that were significant, at a 10% level, in the regression models for at least ten of the thirteen key survey outcomes. The selected interactions employed in the calibration included:

- sex of individual x education of individual
- Black x education of individual
- Black x Hispanic
- Black x region
- Hispanic x education of individual
- Hispanic x region
- Hispanic x FES
- Hispanic x family size
- Hispanic x family unit with foreign-born reference person or foreign-born spouse/partner
- Hispanic x presence of children

- education of individual x FES
- education of individual x family size
- education of individual x presence of children
- region x family size
- region x family unit with foreign-born reference person or foreign-born spouse/partner
- FES x family unit with foreign-born reference person or foreign-born spouse/partner
- family size x presence of children

The calibration adjustment was performed using a raking ratio (or iterative proportional fitting) method (Deming and Stephan, 1940) through a SAS macro developed by Battaglia et al. (2009). An advantage of this SAS macro is that apart from running the raking procedure to adjust the weights to enforce the weighted sample distribution to match the population margins in the selected calibration dimensions, it also simultaneously trims the weights according to trimming parameters, in order to mitigate some potential increase of the variance of statistical estimates due to weight variability.

The final cross-sectional individual weight for the 2023 PSID data was derived from the output weights of this calibration adjustment with trimming. As expected, the PSID weighted estimates of the variables used in the calibration, including age, sex, race, and region, are very close to the weighted estimates based on ACS data; the difference in weighted proportions in each category is less than 0.1%. Appendix Table A2 provides a descriptive summary of the 2023 PSID sample sizes, the distributions of the 2023 cross-sectional individual weight and the ACS population totals in 2023. The variable names for the 2023 Cross-Sectional Individual Weight in the PSID data archive are provided in appendix Table A3.

In 2017, a new immigrant sample was added so the PSID 2019 sample also covers post-1997 immigrants. The cross-sectional individual weight allows analysts to generalize their analysis of individual characteristics to the national population of individuals. For the waves since 2019, the cross-sectional weights attempted to numerically account for all individuals in the United States. However, immigrants arriving after 2017 were not fully covered in the PSID after 2017. The population totals used in calibration include the families comprised of new immigrants and their offspring[4] and, therefore, this procedure also serves as an adjustment to account for the PSID non-coverage of immigrant populations entering the U.S. after 2017.

IV. Recommendations for Using the 2023 PSID Cross-Sectional Individual Weights

[4] This refers to the variable ‘family unit with foreign-born reference person or foreign-born spouse/partner’ used in the calibration.

As indicated in the previous section, the calibration adjustment in the 2023 PSID Cross-Sectional Individual Weights was performed to optimize them with respect to the 13 selected outcome variables. To the extent that the survey variables that users are interested in analyzing are correlated to these set of selected variables, they should expect to obtain similar performances in these estimates, such as bias reductions and improved precision. Therefore, PSID data users are, in general, encouraged to use the cross-sectional weights in any cross-sectional analysis. However, under certain circumstances, weights can be more detrimental – only increasing estimates' variance without reducing any bias, in particular for regression models. For such, users can apply the Fuller (1984) procedure to evaluate if weights are needed for their regression models. What follows is a description of this procedure as laid out in West et al (2025):

1. Fit the model without using the cross-sectional weights but still accounting for the stratification and clustering.
2. Add the cross-sectional weight variable to the unweighted model as an additional predictor, along with its two-way interactions with all other independent variables in the model.
3. Perform a design-adjusted Wald test of the null hypothesis that all of the coefficients associated with these new terms involving the cross-sectional weight are equal to zero.
4. If this null hypothesis is not rejected, the cross-sectional weights are not necessary in the regression model and the unweighted fit can be used.

References

Battaglia, M. P, Hoaglin D. C., and Frankel M. R. 2009. "Practical Considerations in Raking Survey Data." Survey Practice 2 (5). https://doi.org/10.29115/SP-2009-0019.

Chang, W., Nishimura, R., Heeringa, S., Johnson, D., and Sastry N. (2021). "Construction and Evaluation of the 2019 Longitudinal Individual and Family Weights", Panel Study of Income Dynamics Technical Report. Survey Research Center, University of Michigan, Ann Arbor. Available at: http://psidonline.isr.umich.edu.

Chang, W., Nishimura, R., Insolera, N., Sastry N., and Crossley, T. (2025). "Construction and Evaluation of the 2023 Longitudinal Individual and Family Weights", Panel Study of Income Dynamics Technical Report. Survey Research Center, University of Michigan, Ann Arbor. Available at: http://psidonline.isr.umich.edu.

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics, 11(4), 427-444.

Ernst, L.R. (1989). "Weighting issues for longitudinal household and family estimates." In *Panel Surveys* (Eds. D. Kaspryzk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 139-159.

Fuller, W. A. (1984). "Least squares and related analyses for complex survey designs." *Survey Methodology*, *10*(1), 97-118.

Huang, H. (1984). "Obtaining cross-sectional estimates from a longitudinal survey: Experiences of the Income Survey Development Program.", *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 670-675.

Ruggles, S., Flood, S., Sobek, M., Backman, D., Cooper, G., Drew, J., Richards, S., Rodgers, R., Schroeder, J., and Williams, K. IPUMS USA: Version 16.0 [dataset]. Minneapolis, MN: IPUMS, 2025. https://doi.org/10.18128/D010.V16.0.

Johnson, D. S., McGonagle, K. A., Freedman, V. A., & Sastry, N. (2018). Fifty Years of the Panel Study of Income Dynamics: Past, Present, and Future. The ANNALS of the American Academy of Political and Social Science, 680(1), 9–28.

Kalton, G. (1987). "Including nonsample persons in PSID analyses". Panel Study of Income Dynamics Working Paper, Ann Arbor: University of Michigan.

Kalton, G. and Brick, J.M. (1995). "Weighting Schemes for Household Panel Surveys", *Survey Methodology, Vol 21, No. 1,* pp. 33-34, Statistics Canada.

Lavallee, P. (1995). "Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21 25-32.

Little, R.J.A. (1989). "Sampling weights in the PSID: Issues and comments. Panel Study of Income Dynamics Working Paper, Ann Arbor: University of Michigan.

Little, R. J., & Vartivarian, S. (2005). “Does weighting for nonresponse increase the variance of survey means?” *Survey Methodology*, 31(2), 161.

Lynn, P., Buck, N. Burton, J., Laurie, H, Uhrig, S.C.N. (2006). *Quality Profile: British Household Panel Survey, Version 2: Waves 1-13:1991-2003. Essex: University of Essx, Institute for Social and Economic Research.*

U.S. Census Bureau; American Community Survey (ACS), One-Year Public Use Microdata Sample (PUMS), 2023.

West, B. T., Heeringa, S. G., & Berglund, P. A. (2025). *Applied Survey Data Analysis*. CRC Press.

Appendix

Table A1. PSID Size of Sample and Nonsample Individuals and Family Units: 1969-2023

Year	Total Number of Individuals	Total Number of "Sample Persons"	Total Number of "Nonsample Persons"	Total Number of Family Units
1969	17212	16675	537	4460
1970	17349	16359	990	4645
1971	17590	16244	1346	4840
1972	18051	16283	1768	5060
1973	18236	16155	2081	5285
1974	18396	16068	2328	5517
1975	18623	16028	2595	5725
1976	18768	15937	2831	5862
1977	18998	15898	3100	6007
1978	19140	15833	3307	6154
1979	19443	15892	3551	6373
1980	19747	15916	3831	6533
1981	19796	15897	3899	6620
1982	20112	16008	4104	6742
1983	20327	16010	4317	6852
1984	20393	15987	4406	6918
1985	20680	16024	4656	7032
1986	20437	15782	4655	7018
1987	20486	15755	4731	7061
1988	20506	15692	4814	7114
1989	20451	15564	4887	7114
1990	20745	15626	5119	9371
1991	20770	15607	5163	9363
1992	21145	15752	5393	9829
1993	22311	16121	6190	9977
1994	24512	18153	6359	10764
1995	23929	17699	6230	10401
1996	23810	17587	6223	8511
1997	19761	15047	4714	6747
1999	20515	15313	5202	6997
2001	21400	15639	5761	7406
2003	22290	16005	6285	7822
2005	22918	16614	6304	8002
2007	23501	16906	6595	8289
2009	24385	17471	6914	8690
2011	24661	17643	7018	8907
2013	24952	17785	7167	9063
2015	24637	17505	7132	9048
2017	26445	19258	7187	9607
2019	26084	19055	7029	9569
2021	24669	18031	6638	9207
2023	24536	18030	6506	9152

Table A2. Distribution of PSID Cross-Sectional Individual Weights: 1997-2023

Year	PSID						CPS	ACS
	Sample Size	Cross-Sectional Individual Weight					March Supplement Population Total	One Year PUMS Population Total
		Mean	Std Dev	Min	Max	Sum of Weights		
1997	19,761	13,501	10,121	62	68,079	266,792,421	266,792,407	Not Used
1999	20,515	13,246	9,964	32	78,034	271,742,851	271,742,834	
2001	21,400	13,062	10,094	34	76,156	279,517,336	279,517,359	
2003	22,290	12,828	10,099	67	80,408	285,933,473	285,933,409	
2005	22,918	12,705	10,270	69	67,753	291,166,164	291,166,198	
2007	23,501	12,630	10,293	48	68,214	296,824,059	296,824,002	
2009	24,385	12,363	9,311	118	53,258	301,482,827	301,482,827	
2011	24,661	12,413	10,614	66	88,308	306,109,661	306,109,661	
2013	24,952	12,469	10,851	45	85,742	311,116,170	311,116,170	
2015	24,637	13,046	11,756	60	86,506	321,418,821	316,167,949	321,418,821
2017	26,445	12,180	11,415	37	78,618	322,103,607	Not Used	322,103,564
2019	26,084	12,584	12,667	193	52,853	328,239,523		328,239,523
2021	24,669	13,454	13,195	211	56,506	331,893,697		331,893,745
2023	24,536	13,650	13,585	402	57,330	334,914,999		334,914,896

* Due to overlap with the HRS screening for its new cohorts, recent immigrants born between 1960 and 1971 (as well as post-1997 immigrants who co-reside with individuals born in these years) were not part of the PSID NIS-2017 sample. Recent immigrants born between 1960 and 1971 is referenced as the 'donut hole' group. The families in which the reference person and/or the spouse/partner are in the donut hole ('donut hole' families) were not added to the PSID panel until 2019. Individuals living in the donut hole families and individuals who are recent (post-1997) immigrants but live in group quarters were excluded from the ACS estimate in 2017

Table A3. Variable Names for PSID Cross-Sectional Weights

Year	Individual Weight Variable Name	Family Weight Variable Name
1997	ER33438	ER12224
1999	ER33547	ER16519
2001	ER33639	ER20459
2003	ER33742	ER24180
2005	ER33849	Not Computed
2007	ER33951	
2009	ER34046	
2011	ER34155	
2013	ER34269	
2015	ER34414	
2017	ER34651	ER71571
2019	ER34864	ER77632
2021	ER35065	Not Computed
2023	ER35265	