



Technical Paper Series #22-01

Updating the Geospatial Data in the PSID
Restricted Data Enclave for 2005-2017

David Johnson
Mohammad Mushtaq
Nishaad Rao
Noura Insolera
Institute for Social Research, University of Michigan

April 2022

Introduction

The PSID Geospatial Data provide the geographic codes necessary to link individual and family data from the PSID in each year to contextual data from secondary sources such as the Census and American Community Survey (ACS). The PSID Geospatial Data are available in the restricted data enclave and are the most requested restricted files. Over the course of the study, geocoding has been conducted using five different Censuses (see the online [Geospatial File Description](#)). Addresses are geocoded to the relevant Census geographic identifiers (e.g., starting with the 2011 wave, address information has been geocoded to the 2010 census geographic identifiers).

In 2019, a new geocoding procedure was initiated to prioritize the use of the physical address of PSID respondents' residences. During the past several waves (2005-2017), it was determined that for a small subset of cases the U.S. postal mailing address provided by the respondent for the mailing of their interview incentive payment was used to determine the geocode, even when a different physical address was provided earlier in the interview. The current 2019 geocoding process was updated to use the postal mail address only when there is no physical address.¹ The new design for the geocode file (physical address first) will be available for approved users in the enclave and will include the updated address; a previous version (old design, mailing address first) that includes the original addresses will also be available in the enclave for comparison.

This technical note describes the differences in the geocoded locations using this new process (physical address) compared to the previous process (mailing address for some). For the 2005-2017 PSID waves about 3% of families obtained a geocode based on their mailing address. While these families are more likely to move, the overall migration rates using their physical address were not statistically different than those using the mailing address. In addition, the families that previously had a geocoded mailing address have different characteristics than those where a physical address was used. Evaluating the characteristics of the Census tracts where these families live, using either their physical address or the mailing address, does not yield statistically significant different mean estimates for a variety of tract or county characteristics. Overall, the impact on estimates of using this new geocoding procedure is minimal.

Updated geocoded locations are now available in the enclave and will continue to be updated using this new process of using the physical address (as done in the 2019 release). These changes only impact the restricted geocoded files and not the public use PSID data available at www.psid.org. The public use data are regularly updated with corrected data or improved measures (as documented in the [User Guides](#)), and the next release will include a handful of cases where the state of residence has been updated.

Geocoding Process

The geocoding process begins once the family interview is complete. Addresses are updated in order to ensure timely delivery of the respondent payment for completing the interview.

¹ For the 2019 data, the address review and geocoding process was substantially updated and expanded, including additional quality reviews, additional cleaning of street names, cities, states, and zip codes, comparisons to United States Postal Service (USPS) files, and cleaning of abbreviations and misspellings.

Addresses are confirmed using the United States Postal Service (USPS) address system. Currently, the physical and mailing addresses are kept for future reference, and the physical address receives a geocode using the SAS proc geocode and proc ginside procedures. The SAS functions of ZIPCITY and STFIPS are used to validate the zip code and Federal Information Processing Series (FIPS) state code. The physical address is geocoded, and if no physical address is provided, then the mailing address is used. If a P.O. Box address is the only one provided, it is geocoded at the zip code centroid.

Results of New Geocoding Process

Because many researchers use the geocode data to evaluate migration patterns (see Lichter et al. forthcoming) and neighborhood characteristics (see Leibbrand and Crowder 2018), we include comparisons of these items using both the updated (physical) and original (mailing for some) addresses.

Table 1: Number of Addresses that have an updated physical address (from the original mailing address) by whether it is in a different State, County or Tract; 2005-2017

Year	Total addresses	Total updated	Different State	Different County within State	Different Tract within County	Within Tract
2005	8,002	191	48	24	90	29
2007	8,289	178	18	36	93	31
2009	8,690	259	18	40	154	47
2011	8,907	290	40	54	150	46
2013	9,063	298	43	58	144	53
2015	9,048	337	48	60	175	54
2017	9,607	310	39	66	167	38
Total	61,606	1,863	254	338	973	298

Table 1 shows the counts of the changes in geocodes for the 3% of families with an updated address in 2005-2017 (Note: The current 2019 geocoded file available in the enclave already uses this new procedure). Over the seven waves, 1,863 have updated geocodes, which represents 3% of the entire 61,606 families present in these seven waves of PSID data. Most of these families have a different tract within a county. Of these 1,842 families, 1,383 are unique families, suggesting that 25% are families that had a mailing address used in multiple waves. And 68% have their physical address in the same county (with 23% of these geocoded to the same tract).²

Table 2 shows the characteristics of families with an updated physical address. They are more likely to be younger, have lower family incomes, smaller family sizes, and are less likely to be homeowners.

² Some addresses were found missing or in a foreign country and are not included in the analysis.

Table 2: Family and reference person (RP) characteristics for families with updated addresses to those whose address was the same; 2005-2017

Year	Family Income		Female (RP) %		Black (RP) %		Homeowner %		Age (RP)		Family Size	
	Updated	Same	Updated	Same	Updated	Same	Updated	Same	Updated	Same	Updated	Same
2005	64,153	79,349	29.4%	30.0%	25.5%	33.3%	33.3%	60.7%	33.2	45.3	1.90	2.70
2007	51,178	79,286	27.3%	30.3%	14.3%	33.8%	31.8%	59.3%	43.6	45.2	2.00	2.70
2009	56,204	80,338	28.6%	30.7%	19.0%	34.5%	14.3%	56.8%	38.2	45.5	2.00	2.60
2011	69,191	73,701	29.0%	31.8%	19.5%	35.6%	9.8%	54.4%	38.7	45.5	2.00	2.60
2013	64,794	74,567	34.0%	31.9%	27.3%	37.1%	15.9%	52.4%	36.8	45.8	2.10	2.60
2015	64,639	73,361	37.5%	32.7%	12.5%	38.4%	22.9%	51.0%	45.5	46.0	1.70	2.60
2017	64,393	74,056	30.0%	31.8%	30.8%	37.9%	32.5%	51.3%	44.3	46.2	1.90	2.60

Note: Inflation adjusted to 2017 dollars using the Consumer Price Index for All Urban Consumers (CPI-U)

Migration Rates

Figures 1A and 1B show the state-to-state and county-to-county migration rates for families with an updated physical address (updated families) for the period from 2005-2017. As shown, the updated migration rate is higher than the original rate because the continued use of mailing address did not account for the actual moves between physical addresses for these families. The rates of moving for these families are higher than for the average PSID family, with state-level moves at about 13% in 2017, compared to 6.3% overall. This is not surprising since Table 2 shows that these families are younger and more likely to be renters. For most years, the rates for inter-state and inter-county moves are statistically different for these families.

Figure 1A: Percent of families with an updated geocode who move to a different state between waves; 2005-2017 with error bounds

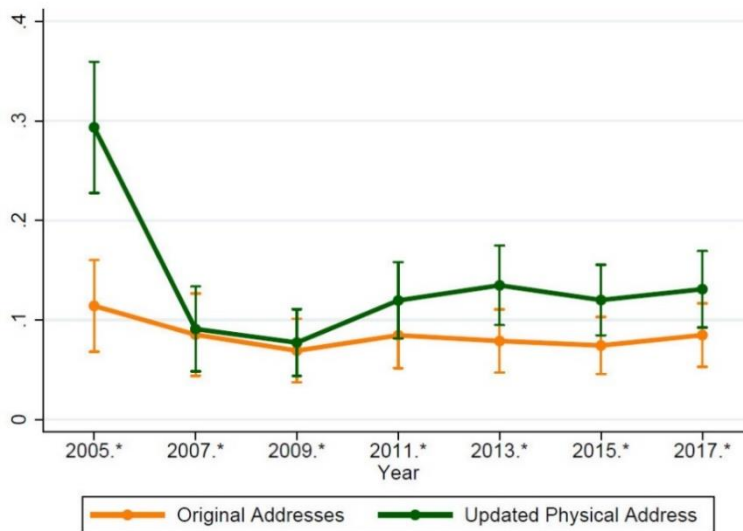
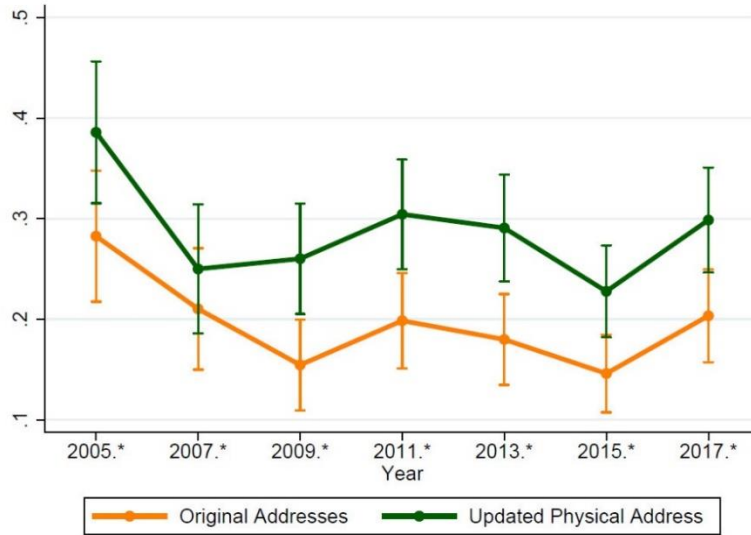


Figure 1B: Percent of families with an updated geocode who move to a different county between waves; 2005-2017 with error bounds



Because these families comprise only a small percentage of all families, the migration rates for the entire sample are also not significantly different. Figures 2A, 2B, and 2C show the overall inter-state, inter-county and inter-tract migration rates for 2005 through 2019. In 2017, the inter-county migration rate (those who moved between 2015 and 2017) is 14.4% in the updated data compared to 13.8% in the original data. While these adjusted figures show higher levels of migration (due to the higher rate of moving for these families), the error bounds demonstrate that these differences are not statistically significant. As shown in other research, and in Table 4, the migration rates in the PSID are more volatile over time than estimates from other data sources (see Johnson and Schulhofer-Wohl 2021).

Figure 2A: Percent of all families who move to a different state between waves; 2005-2019 with error bounds

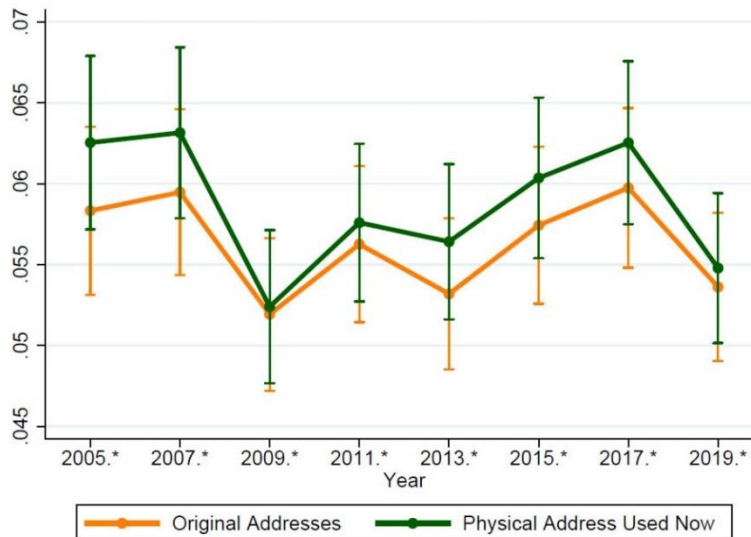


Figure 2B: Percent of all families who move to a different county between waves; 2005-2019 with error bounds

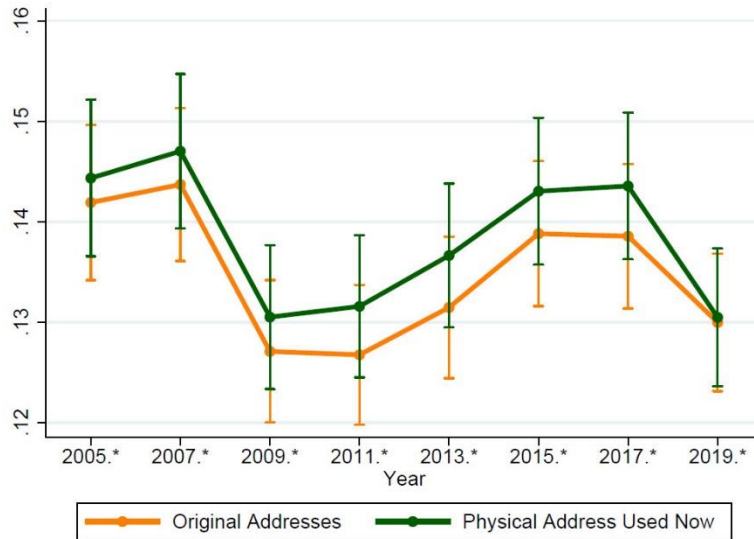
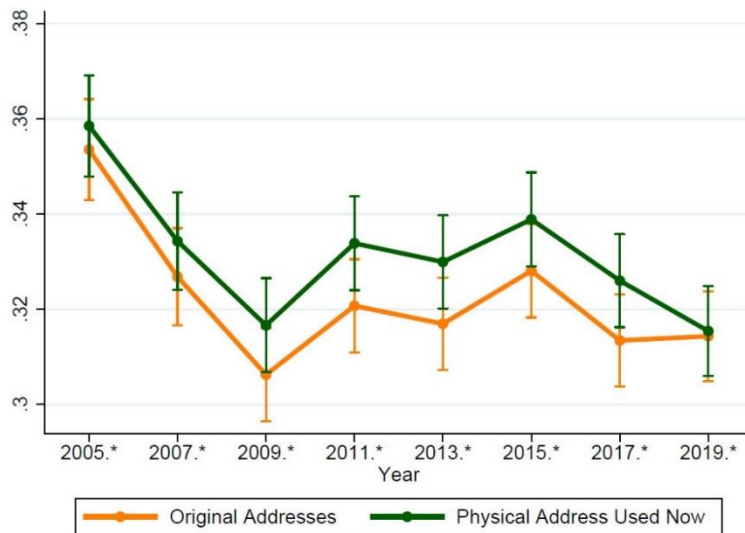


Figure 2C: Percent of all families who move to a different Census tract between waves; 2005-2019 with error bounds



Characteristics of Counties and Tracts for Families

Tables 3A and 3B use ACS data to show the characteristics of counties and tracts for families based on contextual data that have been linked to the PSID data using the geocodes. These tables compare the average characteristics of the counties for the families who have different geocodes (using mailing and physical addresses) for percent female, percent black, percent homeowners, and household income. As can be seen in Table 3B, the updated families live in tracts with lower income and fewer homeowners. As shown in the tables, all estimates across all

years are similar for both the mailing and physical addresses, and none of the differences are statistically different.

Table 3A: ACS Characteristics of the County comparing the original counties (using mailing address) to the updated counties (using physical address)

	Female %		Black %		Homeownership %		Household Income	
	Mailing	Physical	Mailing	Physical	Mailing	Physical	Mailing	Physical
2005	51.2%	51.1%	13.6%	13.8%	64.6%	65.2%	\$ 65,124	\$ 63,245
2007	51.1%	51.1%	15.2%	15.0%	66.4%	66.8%	\$ 71,246	\$ 71,523
2009	50.6%	50.8%	12.8%	15.9%	64.2%	63.9%	\$ 74,053	\$ 71,223
2011	50.9%	51.0%	15.0%	18.2%	61.8%	60.2%	\$ 71,452	\$ 71,498
2013	50.8%	50.9%	15.8%	14.5%	61.9%	59.7%	\$ 75,908	\$ 77,943
2015	51.0%	51.2%	14.2%	14.4%	59.8%	59.2%	\$ 80,680	\$ 83,437
2017	50.8%	50.9%	13.2%	15.4%	62.4%	61.8%	\$ 87,745	\$ 88,405

Household income in 2017 dollars

Table 3B: ACS Characteristics of the Census tract comparing the original counties (using mailing address) to the updated tracts (using physical address)

	Female %		Black %		Homeownership %		Household Income	
	Mailing	Physical	Mailing	Physical	Mailing	Physical	Mailing	Physical
2005	51.9%	50.8%	32.4%	31.4%	21.3%	19.8%	\$ 41,679	\$ 42,665
2007	51.6%	51.9%	35.4%	32.7%	20.2%	21.5%	\$ 37,560	\$ 44,702
2009	51.6%	51.6%	33.7%	32.6%	20.7%	21.1%	\$ 39,615	\$ 44,489
2011	51.7%	52.0%	37.1%	38.2%	21.0%	20.8%	\$ 41,873	\$ 45,335
2013	51.7%	51.7%	37.1%	34.0%	18.3%	21.1%	\$ 36,414	\$ 45,307
2015	51.8%	51.0%	33.1%	32.3%	21.6%	21.7%	\$ 41,640	\$ 45,874
2017	52.1%	51.5%	33.4%	31.9%	21.8%	20.1%	\$ 41,499	\$ 43,349

Household income in 2017 dollars

Comparison to ACS Migration rates

Figure 3 (panels A and B) uses the annual ACS data to show the frequency of households that move across states and counties between the years 2005 to 2017. Since this is an annual measure, the migration rates for inter-state moves are about half of those using the biennial PSID, while ACS inter-county migration rates are about two-thirds of the PSID rates. The annual (ACS) and biennial (PSID) percent of family moves by state, county, MSA and tract are displayed in Table 4. Both the ACS and PSID show a fall in migration rates between 2005 and 2009, a more stable trend from 2009-2015, and a fall since 2017.

Figure 3: ACS annual inter-state and inter-county migration rates, 2005-2019

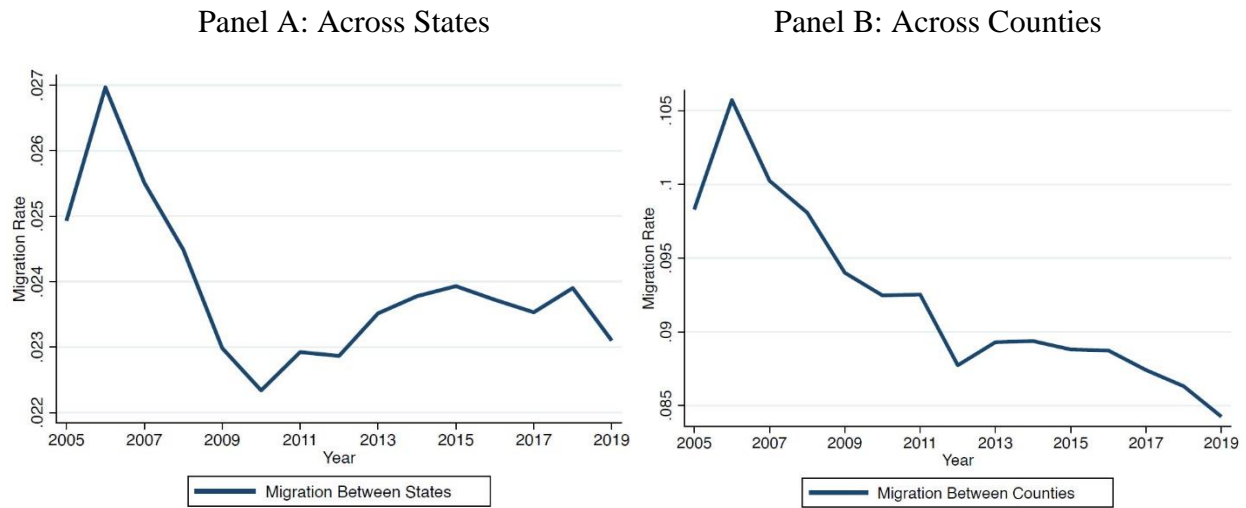


Table 4: Percent of all families who move by State, county, MSA, and tract for PSID and ACS; 2005-2019

	PSID				ACS		
	Across-State	Across-County	Across-MSA	Across-Tract	Across-State	Across-County	Across-MSA
2005	6.25%	14.44%	7.48%	35.85%	2.49%	9.83%	4.29%
2006					2.70%	10.57%	4.87%
2007	6.32%	14.70%	7.86%	33.43%	2.55%	10.03%	4.59%
2008					2.45%	9.81%	4.46%
2009	5.24%	13.05%	7.09%	31.66%	2.30%	9.40%	4.19%
2010					2.23%	9.25%	4.14%
2011	5.76%	13.16%	7.41%	33.39%	2.29%	9.25%	4.22%
2012					2.29%	8.77%	4.59%
2013	5.64%	13.67%	7.79%	32.99%	2.35%	8.93%	4.70%
2014					2.38%	8.94%	4.79%
2015	6.04%	14.31%	8.15%	33.89%	2.39%	8.88%	4.76%
2016					2.37%	8.87%	4.81%
2017	6.25%	14.36%	8.25%	32.60%	2.35%	8.74%	4.75%
2018					2.39%	8.63%	4.73%
2019	5.48%	13.05%	7.43%	31.54%	2.31%	8.43%	4.61%

Conclusion

PSID recently updated its geocoding procedures to use respondents’ physical residential address instead of their mailing address, which affected a relatively small percentage of cases. The change in the address used for geocoding was more likely for families that were younger, lower income, smaller, and renters. This change in address type led to systematic but very small increases in cross-boundary migration rates and minor changes in contextual characteristics to which families were exposed.

References

Johnson, J. and Schulhofer-Wohl, S. "Explaining the Decline in U.S. Internal Migration: The Role of Return Migration," paper presented at PAA 2021.

Leibbrand, Christine and Crowder, Kyle. Migration, Mobility, and Neighborhood Attainment: Using the PSID to Understand the Processes of Racial Stratification. *The ANNALS of the American Academy of Political and Social Science*. 2018. 680, (1): 172-192.
PMCID: PMC6910251

Lichter, Daniel T., Parisi, Domenico, and Taquino, Michael C. Inter-County Migration and the Spatial Concentration of Poverty: Comparing Metro and Nonmetro Patterns. *Rural Sociology*. Forthcoming.