Documentation for the 2013-2017 Employer Geocode Match Files

August 2020

PSID.org
psidhelp@umich.edu

Panel Study of Income Dynamics
Economic Behavior Program
Survey Research Center
Institute for Social Research
University of Michigan
PO Box 1248
Ann Arbor, MI 48106-1248

TABLE OF CONTENTS

2013-2017 Employer Geocode Match Data

The 2013-2017 PSID Employer Geocode Match Files contain location information for the employer of Reference Persons and Spouses/Partners who were currently employed at the time of interview. Address data obtained from respondents were geocoded using the SAS proc geocode and proc ginside processes. Census geocodes on this file are from the 2010 Census. The FIPS county identifiers are current as of the time of release.

This is an "alpha" release. While it has undergone limited editing, reporting or recording errors in the underlying information collected from respondents may lead to errors in the geocoded data (e.g., employer locations that are implausibly distant from the household's place of residence). We discuss some of these issues below.

A more detailed variable-by-variable description of the data is available in the 2013-2017 Employer Geocode Match Files Codebook. The data files are released at the Census Tract, Block-Group, and Block Level.

Data Used for the 2013-2017 Employer Geocode Match Files

The geocoded addresses that are utilized to create these files are from data sets that are kept separately from the main PSID data for the 2013, 2015, and 2017 waves.

In section B/C and section D/E, respondents are asked for the name of the employer of the Reference Person and, if applicable, Spouse/Partner. This information is used to improve question clarity (e.g., to focus respondent on the Reference Person or Spouse/Partner's current employer). Beginning in 2013, both Reference Persons and Spouses/Partners who were working for someone else were asked a sequence of questions to obtain the address where that employer is located and the phone number at that location (see Figure 1). Starting in 2015, this information was asked if the Reference Person or Spouse/Partner was working for a different employer. If the Reference Person/Spouse/Partner was working for the same employer as in the prior wave, we re-asked the information only if the respondent refused to provide this information or provided incomplete information in the prior wave. In cases where we did not need to re- ask the information in current wave, the prior wave report was carried forward.

For the initial processing of this file using 2013 and 2015 waves only we added a special procedure to improve the quality of the year one address collection. For cases with missing information for 2013 and where the individual was working for the same employer and we did re-ask employer information, we filled-in the 2015 information backward to 2013. In some cases, incomplete reports from both waves were combined into a single more complete report that was then attached to both years. Please note this was only done in the initial release of this file using the first two waves of data: 2013 and 2015. As of 2017 we are no longer changing past wave's data using the current year, but rather, we are simply pulling it forward if it was not re-asked in the current wave.

Figure 1: Employer Address Information from Section BC and Section DE

BC19A. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)]
EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])
[*1st ITERATION:* We would like to obtain information about characteristics of employers of our study members, such as benefits offered to employees. In order to do this, we need the full address of the place where [you work / [RP / SP] works] for [EMPNAME].
As you know, the information we collect is confidential, and that includes maintaining confidentiality in collecting information about employers. / *2nd ITERATION:* As we mentioned earlier, we would like to obtain information about employer characteristics, such as benefits offered to employees.
In order to do this, we need the full address of the place where [you work / [SP]] works] for [EMPNAME].
- If needed: We would use the employer address to collect <u>general</u> information about [your / her] employer, such as number of workers or benefits offered to employees.]
- If needed: The information we collect is confidential, and that includes maintaining confidentiality in collecting information about employers.
- If R knows just part of the address, ENTER [1]
- If R knows <u>none</u> of the address, ENTER [Ctrl-D]

| 1. Continue | DK/RF | → GO TO BC50CKPT
↓

BC19ADDR1. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)]
EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])
What is the mailing address?
Address 1
- Enter street address here
- Enter PO Box at next screen
- Ctrl-D and Ctrl-R not allowed; For DK/RF, type [DK] or [REF]

| String 40 |

BC19ADDR2. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)]
EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])
Address 2
- Enter PO Box here
- If none, PRESS [Enter] to continue.
- Ctrl-D and Ctrl-R not allowed; For DK/RF, PRESS [Enter]

| String 40 |

BC19CITY. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)]
EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])
City
- Ctrl-D and Ctrl-R not allowed; For DK/RF, type [DK] or [REF]

| String 40 |

BC19STATE. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)]
EMPLOYER: [EmpName] ([BegMo/BegYr] – [EndMo/EndYr])
State
- Start typing the name of the State to bring up the look-up list
- If foreign country, ENTER [FOR] to select "Foreign Country"

| String 30; U.S. State | | Foreign Country | DK/RF | → GO TO BC19PHONE
↓   ↓

BC19ZIP. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)] EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])	BC19CNTRY. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)] EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])
Zip Code • If foreign country, ENTER all [0]'s • Ctrl-D and Ctrl-R not allowed; For DK/RF, ENTER all [0]'s _ _ _ _ _ - _ _ _ _	Country if different from USA • Start typing the name of the Foreign Country to bring up the look-up list String 30; Country

BC19PHONE. [CYNAMF CYNAML]: [CYAQRTH], [CYFUHU], Age [AGEIWDATE] [(Respondent)]
        EMPLOYER: [EMPNAME] ([BEGMO/BEGYR] – [ENDMO/ENDYR])
        (And) the area code and phone number?
• Area Code and Telephone Number
• If foreign phone, ENTER all [0]'s
• Ctrl-D and Ctrl-R not allowed; For DK/RF ENTER all [0]'s

(_ _ _)_ _ _-_ _ _ _

Data Quality

The quality of the geocoded data depends on the completeness and accuracy of the address information provided by respondents. The SAS geocode process generates several indicators of the precision of the geocoded information. For example, if a respondent provides only the zip code and city of the workplace but not a street address, the geocoding can only be done at the zip code level.

The distributions of match levels for each interview year are reported in Table 1. Nearly 60 percent of the cases were matched at the "dwelling"** level (i.e., exact address) and 80 percent at the zip code level or better.

If a respondent reports an address at the zip code level only (i.e., either doesn't know or refuses to give the street address), the case cannot be matched at the "dwelling" level. More subtly, if a respondent misreports the zip code but correctly reports the city, the result might be a correct match at the city level or an incorrect match at the zip code level, depending on whether the reported zip code is part of the reported city.

For most applications, we recommend that the user compare the geocoded employer data with the household's geocoded residential location. Large discrepancies may signal atypical employment arrangements (e.g., flying to the employer location) but may also be a flag for possible errors in the employer information.

Table 1: Employer Geocode Matches: Match Level by Year					
		Interview Year			
Match Level (RECODED SAS MATCH LEVEL) **		2013	2015	2017	Total
1 = Dwelling** Match	Frequency	4776	4959	5742	15477
	Column %	57.38	59.38	50.70	55.27
2 = Street match - without number or ambiguous house** match	Frequency	589	645	671	1905
	Column %	7.08	7.72	5.92	6.80
3 = Zip match	Frequency	1100	1239	1510	3849
	Column %	13.21	14.84	13.33	13.75
4 = City match	Frequency	445	854	1035	2334
	Column %	5.35	10.23	9.14	8.34
5 = None match	Frequency	198	203	545	946
	Column %	2.38	2.43	4.81	3.38
6 = Not geocoded (too little information or foreign address)	Frequency	1216	451	1823	3490
	Column %	14.61	5.40	16.10	12.46
Total	Frequency	8324	8351	11326	28001
	Column %	100	100	100	100

** We have retained the SAS terminology (dwelling, house), but the data being matched are business addresses, not residences

Geocoding Process

The geocoding process for this file follows the same process used in the geocoding of PSID households' residential addresses. That is, location data were reviewed and cleaned, followed by use of the SAS 9.4 proc geocode process. We imported the latest TIGER/Line shape files for all states from the Census Bureau. These files can be found here: https://www.census.gov/geo/maps-data/data/tiger-line.html.

When an exact match is not found, SAS will use other values to get the closest match. The accuracy of that match is coded in EMPV17 "RECODED SAS MATCH LEVEL". Another accuracy variable, EMPV15, "SAS Numeric Quality of Match", gives a numeric measure of the match's quality. The score is calculated based on EMPV16, "SAS Match Tokens." Each token within EMPV16 has a numeric value that is used to calculate EMPV15. For example, if EMPV16 contains "AD ZC NM" then that means the street name, zip code, and house number matched. The values in EMPV15 equal the sum of 20 for AD, 15 for ZC and 10 for NM for a total score of 45.