

# A Comparison of Person Reported Industry to Employer-Reported Industry in Survey and Administrative Data<sup>1</sup>

Emily Isenberg  
Liana Christin Landivar  
Esther Mezey

U.S. Census Bureau  
4600 Silver Hill Rd.  
Suitland, MD 20746

Axioma, Inc.  
17 State St. Suite 2700  
New York, NY 10004

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

## Abstract

The Census Bureau collects industry information through surveys and administrative data and creates associated public-use statistics. In this paper, we compare person-reported industry in the American Community Survey (ACS) to employer-reported industry from the Quarterly Census of Employment and Wages (QCEW) that is part of the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. This research provides necessary information on the use of administrative data as a supplement to survey data industry information, and the findings will be useful for anyone using industry information from either source. Our project is part of a larger effort to compare information on jobs from household survey data to employer-reported information. This research is the first to compare ACS job data to firm-based administrative data. We find an overall industry sector match rate of 75 percent, and a 61 percent match rate at the 4-digit Census Industry Code (CIC) level. Industry match rates vary by sector and by whether industry sector is classified using ACS or LEHD industry information. The educational services and health care and social assistance sectors have among the highest match rates. The management of companies and enterprises sector has the lowest match rate, using either ACS-reported or LEHD-reported sector. For individuals with imputed industry data, the industry sector match rate is only 14 percent. Our findings suggest that the industry distribution and the sample in a particular industry sector will differ depending on whether ACS or LEHD data are used.

## 1. Introduction

The Census Bureau collects industry information through surveys and administrative data and creates associated public-use statistics. In this paper, we compare person-reported industry in the American Community Survey (ACS) to employer-reported industry from the Quarterly Census of Employment and Wages (QCEW) that is part of the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. This research provides necessary information for investigating the use of administrative data as a supplement to survey data industry information, and the findings will be useful for anyone using industry information from either source. Our project is part of a larger

---

<sup>1</sup> We are grateful for the many helpful suggestions from Melissa Chiu, Mark Kutzbach, and Erika McEntarfer. This paper has not undergone the review accorded Census Bureau publications and no endorsements should be inferred. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

effort to compare information on jobs from household survey data to employer-reported information. This research is the first to compare ACS job information to firm-based administrative data.

The ACS, administered by the Census Bureau, is the largest household survey in the United States. About 3.2 million addresses are sampled each year. The ACS collects household, demographic, and economic data from all residents of a sampled household. Industry data are collected on a person's current, primary job or the most recent job held within the past 5 years if not currently working. The LEHD data covers most of the universe of firms for the United States. The QCEW data includes all firms that are required to report employment for Unemployment Insurance, and the LEHD program has QCEW data for all states.<sup>2</sup> The QCEW employer-level data are linked to quarterly data on individual workers, primarily from Unemployment Insurance (UI) wage records. We link individual work histories from the LEHD program to ACS data using Protected Identification Keys (PIKs) created by the Census Bureau. Because LEHD collects data on all UI covered jobs, individuals may have more than one eligible job around the time of the ACS interview. When there is not a single job match between LEHD-reported job and ACS-reported job, we use ACS interview date and earnings to select a main LEHD job to use for industry comparisons.

We examine the distribution of industry sectors in ACS and LEHD data, compare industry match rates by ACS-reported and LEHD-reported industry sectors, and assess where mismatches occur. We find an industry sector match rate of 75 percent, and a match rate at the 4-digit Census Industry Code (CIC) level of 61 percent.<sup>3</sup> The educational services and health care and social assistance sectors have among the highest match rates. The management of companies and enterprises sector has the lowest match rate, using either ACS-reported or LEHD-reported sector. The wholesale trade sector also has relatively low industry sector match rates. For individuals with imputed industry data, the industry sector match rate is only 14 percent. Those with missing industry are more likely to be young, lower educated, lower income, Black or Hispanic, unemployed or not in the labor force, less likely to have worked in the past year, and worked fewer hours or no hours in the past year.

The only other research we are aware of that compares industry reporting between survey and administrative data is Stinson, Gathright, and Skog (2012), henceforth SGS.<sup>4</sup> They compare industry responses in the Survey of Income and Program Participation (SIPP) to the Census Bureau's Business Register. SGS match individuals in the SIPP to W-2 forms, which contain a firm identifier, and then use the firm identifier to link to businesses in the Business Register.<sup>5</sup> Jobs are matched using employer name and address. Although SGS use different survey and administrative data and a different job-matching algorithm, they find a similar match rate at the 2-digit North American Industry Classification System (NAICS) level of about 75 percent.<sup>6</sup> They also find similar match rates at the 4-digit NAICS level of 63 percent for firms with one establishment and 56 percent for firms with more than one establishment.<sup>7</sup> Wholesale trade and administration and support and waste management sectors also have relatively

---

<sup>2</sup> Massachusetts participates in the LEHD program, but Massachusetts data have not yet been processed.

<sup>3</sup> The Census Bureau has developed and maintained its own industry code list since it started collecting data on industry in 1820. The Census Bureau industry code list has followed the structure of the North American Industry Classification System (NAICS) since its implementation in 1997, but aggregates smaller categories for confidentiality and statistical precision.

<sup>4</sup> SGS also examine firm type (single- or multi-unit), firm size, annual earnings, and class of worker classification. Other research has compared earnings between survey and administrative data, for example, Bound and Krueger (1991), Bound, Brown, Duncan, and Rodgers (1994), Roemer (2002), and Abowd and Stinson (forthcoming). Abraham, Haltiwanger, Sandusky, and Spletzer (forthcoming) compare employment status between survey and administrative data. None of these studies use ACS data.

<sup>5</sup> The Business Register is the Census Bureau's main business list. The Internal Revenue Service (IRS) is the primary source of information on businesses in the Business Register.

<sup>6</sup> Becker et al. (2005), Elvery et al. (2006), and Fairman et al. (2008) discuss differences between the QCEW-based Bureau of Labor Statistics (BLS) business list and the Census Bureau Business Register. The two business lists differ in scope, with the BLS list containing firms required to report earnings for unemployment insurance and the Business Register containing firms reporting to the IRS. Even for businesses in both lists, there can be differences in industry. For example, Elvery et al. (2006) note that of matched single-unit firms, the employment-weighted industry sector match rate is 87 percent.

<sup>7</sup> NAICS sector does not exactly correspond to 2-digit NAICS; some sectors include multiple 2-digit NAICS codes. The 4-digit NAICS codes also do not correspond exactly with 4-digit CIC. Census industry codes follow the

poor match rates in the SGS analyses. To the best of our knowledge, there is no other research examining industry nonresponse and the quality of imputed industry in surveys.

Our findings suggest that the industry distribution and the sample in a particular industry sector will differ depending on whether ACS or LEHD data are used. This variability is relevant for anyone using ACS or LEHD public-use statistics and anyone using industry variables in ACS or LEHD microdata. For example, certain industry sectors, such as construction and manufacturing, tend to be more affected by recessions, and research that focuses on those sectors may have different results depending on whether survey or administrative data are used. Researchers with access to matched data may want to test the sensitivity of their results to different industry variables. Researchers may also want to investigate how industry sector samples differ, for example, how the manufacturing sector differs between ACS and LEHD data. The low match rate for imputed industry suggests that ACS microdata users should use imputed industry values with caution.

## 2. Data

### 2.1. American Community Survey Data

The American Community Survey is a mandatory household survey conducted by the U.S. Census Bureau. It was created to replace the decennial census long form and to provide information on the U.S. population on an annual basis instead of every ten years. The questions in ACS are based on the former Census long form. The ACS was fully implemented in 2005 and in 2006, the ACS was expanded to include group quarters, so that geographic areas could be fully described.<sup>8</sup> The survey is an ongoing survey based on monthly samples, with microdata and public use statistics consolidated for each year.

We use the 2009 ACS population file for our analyses.<sup>9</sup> Our analyses require the use of internal Census ACS data. To meet requirements for confidentiality protection, the Census Bureau implements a variety of strategies including “data swapping.” Data swapping is a method of disclosure avoidance done by editing the source of the data or exchanging the records of a sample of cases.<sup>10</sup> The marginal totals for an area are not affected, but the responses for any given individual on the file may have been swapped with the responses of another similar individual. The file we use for analyses contains pre-swapped, edited data, i.e. variable values have not been swapped for confidentiality protection, but variables have been edited and imputed. In some tables, as noted, we also use unedited variables from this file. We require the use of pre-swapped files to be able to match the response date for the survey and the reported industry to the appropriate LEHD reporting quarter.

The ACS contains a series of questions about household relationships, demographics, income and benefits, and where individuals live and work. The survey questions of most relevance to this project are questions 29, 35 to 38, and 41 to 44, which ask about employment status, type of employer, and industry.<sup>11</sup> See Figure 1 for questions 29, 35 to 38, and 41 to 44 of the survey. We use information on employment status and type of employer to select individuals likely to be in LEHD data, described in more detail in Section 3. We use the Employment Status Recode (ESR) variable to select those employed and at work in the past week. The ESR variable has 6 categories: 1) employed, at work; 2) employed, not at work during the reference week; 3) unemployed; 4) military, at work; 5)

---

structure of NAICS but differ in level of detail. CIC level of detail varies for two primary reasons: Census Bureau coders must be able to obtain enough detail from respondents to code a response to an industry code and an industry must be large enough to meet disclosure restrictions. Industries are aggregated based on the NAICS hierarchy when there is insufficient detail or it is too small to be reported separately.

<sup>8</sup> The *Design and Methodology Report: American Community Survey* (2009) contains information on ACS program history in Chapter 2. The report also contains much information on survey design and methodology (as the title suggests). [http://www.census.gov/acs/www/Downloads/survey\\_methodology/acs\\_design\\_methodology.pdf](http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf)

<sup>9</sup> At the time we started this project, 2009 ACS data was the most recent available. Since then, 2010 and 2011 ACS data have become available. However, the 2010 ACS had a drop in response rate, presumably due to the 2010 decennial census, so we focus on 2009.

<sup>10</sup> Data swapping is discussed in section 13.6 of *The Design and Methodology Report: American Community Survey* (2009).

<sup>11</sup> PDF copies of the ACS questionnaire can be found on the ACS website: [http://www.census.gov/acs/www/methodology/questionnaire\\_archive/](http://www.census.gov/acs/www/methodology/questionnaire_archive/).

military, not at work during the reference week; and 6) not in the labor force.<sup>12</sup> For our analyses, we select individuals who are employed, at work.

The ACS class of worker (COW) variable is based on question 41 which asks about type of employer: private (for profit/not-for-profit), government (local/state/Federal), self-employed (not incorporated/incorporated), or working without pay in a family business or farm. Individuals respond to the questions about type of employer and industry if they worked in the last 5 years.<sup>13</sup> The survey requests information for the primary job if currently working.<sup>14</sup> If the person is not currently working, he or she reports the most recent job held within the past 5 years. From the ACS questionnaire for questions 41-46:

41-46 Current or most recent job activity. Describe clearly this person's chief job activity or business last week. If this person had more than one job, describe the one at which this person worked the most hours. If this person had no job or business last week, give information on his/her last job or business.

Question 42 asks employer name to assist with industry classification. In question 43, the person is asked to describe the industry activity at the location where employed; the industry question is open-ended. Question 44 asks whether this activity was mainly manufacturing, wholesale trade, retail trade, or something else.<sup>15</sup> The Census Bureau uses questions 42 through 44 to code the responses into 269 4-digit Census Industry Code (CIC) categories.<sup>16</sup> The CIC categories are based on the 6-digit 2007 North American Industry Classification System (NAICS). The Census Industry Codes are generally less detailed than the NAICS codes because of potential disclosure risk, but are designed so that 6-digit NAICS codes can be mapped to 4-digit CIC codes.

ACS coding takes place in a centralized location with standardized procedures and a dedicated staff of fully-trained coders. Clerical staff at the Census Bureau's National Processing Center convert the written questionnaire responses to Census Industry Codes using the [Alphabetical Index of Industries and Occupations](#). Clerical coders receive extensive training and must maintain a 94 percent coding accuracy rate to remain qualified for coding. Industry coding is independently verified by additional qualified coders as part of the quality assurance process. Following industry coding, the Census Bureau checks for consistency between industry and other related variables, such as occupation and type of employer. If industry is missing after editing, industry is imputed using information from a "similar" person.<sup>17</sup> We use the edited and imputed industry variable, IND, for most of our analyses.

ACS internal-use data includes an interview or reference date variable, RDATE. Interview date is critical for linking to the appropriate quarter of LEHD data. The ACS questionnaire is mailed to sample addresses. If the household does not mail back the survey, the Census Bureau follows up, first with an attempted telephone interview, and, if

---

<sup>12</sup> ESR is based primarily on questions 26, 29, and 35 to 37 of the survey. Question 26 asks whether the person has ever served in the military and if he or she is now on active duty. Question 29 asks whether the person worked last week. If the person did not work last week, he or she is requested to respond to questions 35 to 37. Those questions ask if the person was on layoff or temporarily absent from a job, if the person was actively looking for work (if not temporarily absent from a job), and if the person could have started a job if offered one (if not temporarily absent from a job). Question 38 asks when the person last worked: within the past 12 months, 1 to 5 years ago, or over 5 years ago or never worked.

<sup>13</sup> For questions 41-46, the questionnaire states: "Answer questions 41-46 if this person worked in the last 5 years. Otherwise, skip to question 47."

<sup>14</sup> If the respondent has more than one job and cannot select a primary job, he or she is prompted to select the job at which he or she worked the most hours in the last week.

<sup>15</sup> Other surveys that ask about industry of place of employment, such as the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP), also ask open-ended questions about industry. The ACS, CPS, and SIPP all have follow-up questions about whether the employer is mainly manufacturing, wholesale trade, retail trade, or something else. The SIPP also includes service in the follow-up question.

<sup>16</sup> Through 2011, CIC was 100 percent coded by clerical staff at the National Processing Center in Jeffersville, Indiana. In 2012, ACS developed an industry auto-coder that used model-based coding for some industry values.

<sup>17</sup> *American Community Survey, Puerto Rico Community Survey, 2009 Subject Definitions* (2009), pp. 84-86, contains information on how industry is coded, edited, and imputed.

[http://www.census.gov/acs/www/Downloads/data\\_documentation/SubjectDefinitions/2009\\_ACSSubjectDefinitions.pdf](http://www.census.gov/acs/www/Downloads/data_documentation/SubjectDefinitions/2009_ACSSubjectDefinitions.pdf)

that does not succeed, with an attempted in-person interview. In completed mail returns, RDATE is the date the respondent prints on the cover of the questionnaire as “today’s date.” In computer-assisted telephone interviews (CATI) or computer-assisted personal interviews (CAPI), RDATE is the last date on which data were collected. An ACS CATI or CAPI interview is intended to be completed during a particular calendar month. If the ACS questionnaire is completed over multiple interviews, RDATE could be at most 30 days after the initial response to interview questions.<sup>18</sup> We expect most mail surveys or interviews are completed on the same day they were started.

## 2.2. LEHD Data

We use employer-reported industry collected as part of the Census Bureau’s Longitudinal Employer-Household Dynamics (LEHD) program.<sup>19</sup> LEHD data are based on worker-level and employer-level administrative data. Individual-level earnings data are derived from state administrative records of the Unemployment Insurance (UI) system. They contain quarterly earnings, by job, for individuals employed in a particular state. An individual may have earnings from more than one employer in a particular quarter, and a job is a particular individual-employer combination. Quarterly earnings are based on when individuals were paid, not when the work was performed. For example, if an individual was paid in quarter  $t$  for two weeks of work, where one week of work was in quarter  $t$  and the other week of work was in the previous quarter, we would only observe earnings in quarter  $t$ .

For all states except Minnesota, earnings are reported at the employer level, by State Employer Identification Numbers (SEIN).<sup>20</sup> Employers may have multiple establishments. Establishments are operating units of the employer and each establishment usually corresponds to a particular location of business. One example of multiple establishments would be several branches of a bank. In Minnesota, we know in which establishment an individual worked. For all other states, we know the employer but we are not able to link an individual to a particular establishment if the employer has more than one establishment.

In addition to the individual-level data, the LEHD program has employer-level administrative data derived from the Quarterly Census of Employment and Wages (QCEW).<sup>21</sup> The QCEW contains information on employer industry and location – state, county, and street address. QCEW data are derived from the quarterly tax reports submitted by employers subject to state Unemployment Insurance laws. Each state has an office or department that compiles and processes the data. For each state, a new business fills out an initial form that determines liability for state Unemployment Insurance (UI) and asks for address and industry information. This form is called the Status Determination Form (SDF). After determining UI liability, the firm reports monthly employment and earnings each quarter. Every three, four, or five years, depending on funding availability, firms file an Annual Refiling Survey, which verifies location of the establishment and industry.

The specific forms used to determine initial UI liability vary by state. In Virginia, for example, firms are asked an open-ended question about industry: “Describe the kind of business in Virginia, giving specific details of items, customers, etc., such as retail-women’s clothes; wholesale-office equipment; construction-single family home, etc.”<sup>22</sup> In the QCEW, industry is reported using 6-digit NAICS. State QCEW offices use information from the Status Determination Form, state and local knowledge of the establishment, and/or contact the company directly to assign a 6-digit NAICS code. If needed, a state will mail a Non-Classified Account (NCA) form to the establishment. The NCA asks for more information about the establishment, requesting a description of the activities and to provide an approximate percentage of sales or revenues from each activity. Individuals in the state QCEW offices receive in-person NAICS coding training from the Bureau of Labor Statistics (BLS), which oversees the QCEW program. The

---

<sup>18</sup> Based on correspondence with Todd Hughes, Assistant Division Chief of Data Collection at the Census Bureau.

<sup>19</sup> Abowd et al (2005) provide detailed information about LEHD data.

<sup>20</sup> The employer, or SEIN, is equivalent to a firm if all parts of the firm operate in one state. For firms that have operations in multiple states, the SEIN would not be equivalent to a firm identifier.

<sup>21</sup> The [Quarterly Census of Employment and Wages \(QCEW\)](http://www.bls.gov/qcew/) program at the Bureau of Labor Statistics publishes a quarterly count of employment and wages reported by employers covering 98 percent of U.S. jobs, available at the county, metro area, state, and national levels by industry.

<sup>22</sup> The Virginia Report to Determine Liability for State Unemployment Tax (VEC FC-27) is available online: [http://www.vec.virginia.gov/vecportal/employer/pdf/vec\\_fc\\_27new.pdf](http://www.vec.virginia.gov/vecportal/employer/pdf/vec_fc_27new.pdf). Instructions are here: [http://www.vec.virginia.gov/pdf/vec\\_fc\\_27ins.pdf](http://www.vec.virginia.gov/pdf/vec_fc_27ins.pdf). Information about the form can be found at: <http://www.vec.virginia.gov/employers/tax-registration>.

industry coders use a software application, AutoNAICS, to assist in coding. AutoNAICS contains descriptions of NAICS codes from the NAICS codebook and enables key word searches. Regional BLS offices also review a sample of NAICS codes assigned by state staff.<sup>23</sup> A firm can request a NAICS code change if the firm thinks the assigned NAICS code is inaccurate.<sup>24</sup>

In some states, NAICS codes are used to determine Unemployment Insurance tax rates, at least for the first several years a firm is in operation, after which the firm's UI claim history is used to determine rates. In those states, QCEW offices are required to obtain NAICS codes for all firms. In addition, the Annual Refiling Survey (ARS), where firms verify industry and location information, is mandatory in some states. In other states, completing the ARS is not mandatory, but states must obtain certain response rates, such as 75 percent for Washington State, in order to comply with states' QCEW contracts with BLS.

The QCEW data received by the LEHD program may have missing NAICS codes. The LEHD program uses longitudinal information from establishments and employers to fill in missing industry information. For employers with no longitudinal information, industry is imputed. Our analysis data has completed LEHD industry information. It is difficult to calculate an overall industry missing rate for the QCEW data used as an input for our analysis data. However, for a selection of three states and two different quarters of data, establishment industry was missing between less than one percent and 14 percent of the time. Between less than 1 percent and 16 percent of employers were missing NAICS information for all establishments; many employers have only one establishment. As noted, in states where QCEW NAICS codes are used to determine UI tax rates, we would expect essentially zero missing rates.

The individual-level UI data is linked to the employer-level QCEW data by SEIN. For employers with more than one establishment in states other than Minnesota, the LEHD program multiply imputes establishments for individuals. Location will likely vary across establishments and industry may vary across establishments. UI and QCEW data are collected by each state and then shared with the LEHD program as part of the Local Employment Dynamics (LED) Partnership. The LEHD program currently has data from all states, including the District of Columbia. Massachusetts recently joined the LEHD program and Massachusetts data have not yet been processed. In these analyses, we use LEHD data from all states except Massachusetts.

A limitation of the LEHD data is that not all employment is covered by the state UI systems. The areas not covered include some agricultural employment, independent contracting, self-employment, military employment, federal civilian employment, railroad employment, some elected state and local government officials, and postsecondary work-study students.<sup>25</sup>

### **2.3 Sample Selection and Link to Census Protected Identification Keys (PIKs)**

From the 2009 ACS data, we select a sample of individuals who are most likely to match to a job in the LEHD data. We keep individuals who are employed, at work, and not in the Armed Forces (ESR=1). We focus on individuals who are currently employed so that we can use ACS interview date (RDATE) to link to quarter of employment in the LEHD data. We exclude federal government employees (COW=5), self-employed not incorporated (COW=6), self-employed incorporated (COW=7), and unpaid family workers (COW=8). Since LEHD data does not currently include Massachusetts, we exclude individuals who report working in Massachusetts (POWS=025).

In the LEHD data, individuals are identified by Protected Identification Keys (PIKS), which are assigned by the Census Bureau based on personal identifying information in the UI data. In order to match ACS and LEHD data, the ACS data also needs PIKS. After selecting a subsample of the 2009 ACS data, we link the ACS data to PIKS using a crosswalk. Like with the LEHD data, the Census Bureau uses personal identifying information on the ACS survey to assign PIKS. The ACS crosswalk file links ACS identifiers to the PIKS. In a small number of cases, less than 0.2 percent, more than one ACS individual links to the same PIK. In those cases, we select the ACS individual that

---

<sup>23</sup> Information on assignment of 6-digit NAICS codes provided by Amanda Chadwick at the Bureau of Labor Statistics and Molly Webster at the Washington State Employment Security Department.

<sup>24</sup> Request for NAICS code change in California: <http://www.labormarketinfo.edd.ca.gov/Content.asp?pageid=176>

<sup>25</sup> The LEHD program is in the process of adding federal workers and the self-employed. For detailed information on UI covered employment see Stevens (2007).

looks most similar to the LEHD individual, based on where in the Census PIK assignment process an individual gets a PIK and PIK match scores.<sup>26</sup> About 92 percent of individuals in our ACS subsample match to PIKs. We use this group for our subsequent analyses.

### 3. Job Match Algorithm

#### 3.1. LEHD Main Job Selection

We link the PIKed ACS subsample to LEHD earnings history data and then, using a set of rules described below, select the LEHD job that is most likely to correspond to the job referenced in the ACS survey. For the LEHD earnings history data, we use a custom national person history file (PHF\_B) and include earnings through 2010 quarter one. We refer to the LEHD job selected as the LEHD main job and it is an SEIN-year-quarter observation. We use ACS interview date (RDATE) and LEHD quarterly earnings to select the LEHD main job.

There are several challenges associated with selecting an LEHD job (PIK-SEIN in a particular quarter) to correspond with the ACS job. First, two timing issues complicate selecting the appropriate quarter for the LEHD main job. First, the ACS survey asks to describe job characteristics for last week's job, which means that employment status and job characteristics match a job approximately one week prior to RDATE. Second, for ACS CATI and CAPI interviews that took place over multiple days, RDATE may not be the same as the day that the person responded to interview questions on job characteristics.<sup>27</sup> (For mail responses, RDATE is respondent reported "today's date" and any follow-up CATI or CAPI interviews use RDATE to form the reference period for interview questions. A respondent may also start filling out a mail questionnaire on one day and complete the questionnaire on a different day. In that case, we would not know whether they used RDATE as the reference date for all questions, as directed.) As noted above, RDATE is at most 30 days within the same calendar month before an initial CATI or CAPI interview. Both of these timing issues mean that in some cases we should select an LEHD main job from the same quarter as RDATE, while in other cases the LEHD main job may be from the quarter before the RDATE quarter.<sup>28</sup> For our job match algorithm, if RDATE is in the last two months of the quarter, we select as potential LEHD main jobs all SEINs in the RDATE quarter. If RDATE is in the first month of the quarter, we select as potential LEHD main jobs all SEINs in the RDATE quarter and all SEINs in the previous quarter. Figure 2 has a flow chart of the job match process.

A second job-match challenge is that after selecting one or two quarters that could correspond to the ACS job, a person can match to more than one job (SEIN). The person may have multiple SEINs in the same quarter or multiple SEINs across the two possible quarters. Multiple SEINs exist in the same quarter because an individual changed jobs in that quarter (worked at different jobs at different times) or an individual worked at multiple jobs at the same time. For ACS survey respondents with more than one job last week, the survey asks the respondent to reference the primary job or the job with the most hours worked. For most states, LEHD data does not contain information on hours worked.<sup>29</sup> For our job match algorithm, we have two approaches for selecting one LEHD main job (SEIN), depending on whether RDATE is in the first month of the quarter. If RDATE is in the last two months of the quarter, we selected potential LEHD main jobs from LEHD jobs in the RDATE quarter. If there is one potential job,

---

<sup>26</sup> In a few cases, it is not possible to tell which ACS individual is more similar to the LEHD individual based on information from the PIK process. In those cases, we arbitrarily select an ACS individual.

<sup>27</sup> An ACS interview could also take place over multiple days if the respondent did not have time to complete a CATI or CAPI interview in one sitting. For a 2009 ACS subsample based on our sample restrictions, 69 percent of respondents responded by mail, 12 percent by CATI, and 18 percent by CAPI. The subsample included individuals who were employed, at work (ESR=1), worked for private, state, or local government employers (COW not equal to 5, 6, 7, or 8) and who worked in a state other than Massachusetts (POWS not equal to 025). We do not know what share of CATI/CAPI interviews took place over multiple days.

<sup>28</sup> It is also possible that someone becomes employed in the last week or two of a quarter but does not get paid until the following quarter. In this case, they may report a job in the ACS that corresponds to a job in the LEHD data from the following quarter. We are not including LEHD jobs in the quarter following RDATE.

<sup>29</sup> Minnesota and Washington State provided LEHD with data on hours worked.

we keep that job (SEIN) as the LEHD main job. If there is more than one potential job, we select the job (SEIN) with the highest earnings.<sup>30</sup> We expect this to correspond with the job with the most hours worked in most cases.

If RDATE is in the first month of the quarter, plausible LEHD main jobs may be in the RDATE quarter or in the previous quarter. If there is one plausible LEHD main job, we select that job as the LEHD main job. If there is more than one plausible LEHD main job and all plausible jobs are in the same quarter, we select the job with the highest earnings. If there is more than one plausible LEHD main job and there are plausible jobs in both the RDATE quarter and the previous quarter, we use more detailed RDATE information. We assume that all survey information was collected on RDATE, ignoring any potential multi-day surveys. If RDATE is in the first seven days of the quarter, we select jobs from the quarter before RDATE and, if there is more than one job, we select the job with the highest earnings. If RDATE is after the first seven days of the quarter, we select jobs from the RDATE quarter, and if there is more than one job, we select the job with the highest earnings.<sup>31</sup> After selecting LEHD main jobs, each individual has one LEHD main job and we use this sample for our industry comparisons.<sup>32</sup> Recall that an illustration of the job matching process is found in Figure 2.

The third challenge with selecting an LEHD main job for an industry comparison is that a SEIN may have multiple establishments (SEINUNITs), and the establishments may have different industries. For most of our analyses we use employment modal industry for each SEIN and do not use SEINUNIT level industry. In our analysis sample, 42 percent of LEHD employers (SEINs) are multi unit and 58 percent of employers are single unit. Of multi units, 61 percent have 4-digit CIC variation across units, which means that of all employers, 26 percent are multi units with industry variation across units. Of multi units with industry variation, an average of 76 percent of employment for the employer is at units with the employment modal CIC. Therefore, we would expect approximately 6 percent individuals in our analysis sample to have a unit level 4-digit CIC that differs from the employment modal 4-digit CIC.<sup>33</sup> Some units that have a different 4-digit CIC from the employment modal CIC may still be in the same industry sector. Multi unit employers are discussed further in the Industry Comparison section.

For our analysis sample, we convert the LEHD SEIN employment mode 6-digit NAICS code to the corresponding 4-digit CIC code. Based on the 4-digit CIC codes for the ACS and LEHD industry variables, we create ACS and LEHD NAICS sector variables (with 20 NAICS sectors). For a full list of the 20 sectors, see Table 1. Some of our comparisons and industry distributions use sector-level industry information. We do not use ACS weights for our analyses. The pre-swapped edited file we use does not contain a person weight. The Appendix discusses how weights may affect the ACS industry distribution.

Table 2 lists observation counts for sample selection and job matching. The 2009 ACS population file has 4.5 million observations, where each observation is an individual. The largest group excluded from our analyses is individuals not employed, at work (ESR not equal to 1), 2.5 million observations. This group includes individuals age 15 and under who are not eligible for the employed, at work universe. Class of worker exclusions and working in Massachusetts affect a much smaller number of individuals. (Excluded groups are not mutually exclusive.) After ACS sample restrictions, we have 1.7 million observations, and after merging to the ACS-PIK crosswalk, we have 1.6 million observations with PIKs. About 99 percent of the PIKed ACS sample merges to the LEHD earnings history file (PHF\_B). The earnings history file includes earnings from 1990 quarter one through 2011 quarter four, as available for each state. However, when we limit the earnings history file to jobs with positive earnings in the ACS interview quarter or the previous quarter, 92 percent of the PIKed ACS merge to the LEHD earnings history

---

<sup>30</sup> In a small number of cases, 0.04 percent, there were multiple SEINs with identical quarterly earnings for the individual. This appeared to be a reporting issue at least some of the time; employers would have identical or very similar earnings histories for the individual. We drop individuals with non-unique maximum earnings across employers.

<sup>31</sup> As when RDATE is in the last two months of the quarter, we drop individuals with non-unique maximum earnings across employers after selecting employers from the acceptable quarters.

<sup>32</sup> It is possible that some LEHD main jobs do not in fact correspond to the ACS job. This could happen due to the complexities of matching noted in the text or if an individual reported an ACS job that is not included in UI earnings records. An alternative job matching approach could include employer name and potentially address matching.

<sup>33</sup> In our analysis sample, each individual is associated with one employer. An average of 24 percent of employment is not at the employment modal 4-digit CIC. 26 percent of employers that are multi unit with CIC variation across units. (24 percent of 26 percent = 6 percent.)



file; we now have a sample with 1.4 million observations. We further select quarters that have plausible LEHD main jobs, based on ACS interview date (RDATE) as described above. Of individuals with at least one plausible LEHD main job, 87 percent have one plausible main job and 13 percent have multiple plausible main jobs.

### 3.2. Job Match Assessment

Table 3 compares place of work state, county, and tract information between the LEHD data and ACS data. As a robustness check, we are interested in whether state/county/tract match better in the jobs that we have selected as main jobs compared to alternative plausible jobs. We expect some differences in geography, especially at the county and tract level, since the ACS asks where someone reports to work, while LEHD data has the address for the employer. This may be problematic for contractors, who work at one employer but are paid by another employer, individuals who work at temporary help agencies, and construction workers, to name a few examples. LEHD data also has a known address-reporting issue for school districts. Many school districts report only one UI address for all workers, usually the school district headquarters, and do not report employment by individual school. However, we expect main jobs to have better geographic match rates than alternative plausible jobs. This provides some evidence of whether we are selecting the correct main job from our job-match algorithm. Main jobs do appear to have substantially higher geography match rates than alternative plausible jobs.<sup>34</sup>

Our data has very high state match rates – 98 percent for main jobs, 89 percent for other plausible jobs. In LEHD data, state is at the SEIN level, while county and tract place of work information is at the SEINUNIT level. For single units, we know the SEINUNIT and associated geography information. SEINUNIT is imputed for multi units and we use geography information from the first implicate of the SEINUNIT impute.<sup>35</sup> Match rates are lower for county and tract. For main jobs, 73 percent have matching county information, and only 47 percent have matching tract information. For alternative plausible jobs, 51 percent have matching county information and only 13 percent have matching tract information.<sup>36</sup> We may find a lower match rate for main jobs for individuals with more than one plausible job, since there could be timing errors in identifying the job worked at the time of the ACS interview and, in the case of multiple jobs held at the same time, ACS respondents describe the job with the most hours worked last week and we use earnings to select the main job. The match rates for main jobs for PIKs with more than one plausible job are lower than the match rate for all main jobs, which includes many cases with only one plausible job, but are still higher than the match rate for alternative plausible jobs. The relatively low tract match rates suggest that there could be some challenges in using address to select main jobs, since county and tract information disagree in a substantial number of cases and there is a high ACS missing rate for tract information.

In Minnesota, SEINUNITs are reported for workers, and we calculate similar county and tract match rates using only Minnesota. The county and tract match rates are higher for all main jobs and main jobs for PIKs with more than one plausible job. This suggests that the lack of SEINUNIT information for workers in states other than Minnesota is contributing to the lower county and tract match rates.<sup>37</sup>

In other analyses not shown, we find that place of work geography match rates vary across industry sectors. We would expect lower geography match rates in sectors where physical place of work differs from address of employer, such as the sector with temporary help agencies – administrative and support and waste management. The sector does indeed have lower tract and county match rates than most other sectors. Mining, quarrying, and oil and gas extraction has the lowest tract match rates; the physical place of work may often differ from address of employer.

---

<sup>34</sup> We do not distinguish by LEHD geography data quality for these analyses.

<sup>35</sup> We have SEINUNIT matched to individual workers for Minnesota.

<sup>36</sup> The ACS place of work information may be edited, imputed, or missing. We would expect worse geography matches when ACS place of work is edited or imputed. One percent of jobs in each group are missing ACS county information; 21 percent of main jobs are missing ACS tract information; 23 percent of main jobs with more than one plausible job and 23 percent of alternative plausible jobs are missing tract information.

<sup>37</sup> The Census Bureau has an ongoing project that will examine place of work geography matching between the ACS and the QCEW in more depth than this paper.

#### 4. Industry Comparison

There are a number of possible reasons for ACS and LEHD industry data mismatches. The ACS survey respondent, responding for him- or herself, may misunderstand the work done by his or her employer, or the ACS survey respondent responding for another household member may not be sufficiently familiar with the work of the employer of the other household member. Both the ACS and LEHD data ask open-ended industry questions, and the text is then coded to 4-digit CIC or 6-digit NAICS codes. The industry questions for ACS and LEHD are worded differently and people may respond with different information. The coding procedures for the ACS and LEHD data may also be different, resulting in different numeric codes for similar text.<sup>38</sup> To the authors' knowledge, there has been no systematic study comparing the industry coding and editing procedures of the ACS to the industry coding and editing procedures of the LEHD, or other employer-reported data. Finally, for most of our analyses we use SEIN employment mode industry from the LEHD data. If the employer has multiple establishments with industry variation across establishments, the individual could actually be working at an establishment with a different industry. We do not expect this to be a large issue, given the multi unit discussion in the Job Match Algorithm section (Section 3).

Table 4 presents industry match rates for main jobs at the sector level and the 4-digit CIC level. Using the edited ACS industry variable (IND) and the LEHD SEIN employment mode industry, there is an industry sector match rate of 75 percent, and a match rate at the 4-digit CIC level of 61 percent. We separately compare the industry match for when industry is as reported, assigned, or imputed in the ACS.<sup>39</sup> The match rate is substantially better for the as-reported cases. ACS industry is imputed in three percent of our analysis sample and assigned in 0.1 percent of our analysis sample. The industry sector match rate drops from 77 percent for as-reported cases, to 65 percent for assigned cases, down to 14 percent for imputed cases. Match rates similarly decline for 4-digit CIC.

We also assess whether industry match rate may differ for LEHD single unit and multi unit employers. For multi unit employers, some workers may be working at a SEINUNIT with a different industry from the SEIN employment mode industry. Of main job employers, 58 percent are single unit; 42 percent are multi unit. Industry match rate is similar for single unit and multi unit SEIN; 74 percent for single units and 77 percent for multi units. Surprisingly, the match rate is slightly higher for multi units. Match rates between single and multi units are also similar at the 4-digit CIC level, with multi unit SEIN having a slightly higher match rate. In the second panel, we calculated industry match rates using SEINUNIT industry. For multi units, we use the SEINUNIT from the first implicate of the SEINUNIT impute. The match rates using SEINUNIT industry are slightly lower than industry match rates using SEIN employment mode industry; 73 percent match for industry sector and 58 percent match for 4-digit CIC. Match rates are similar for single units and multi units, with slightly higher match rates for single units.

Figure 3 compares the industry sector distribution for the ACS and LEHD industry variables (IND and SEIN employment mode industry). In figures, ACS sector refers to the ACS IND sector and LEHD sector refers to the LEHD SEIN employment modal sector. (The distributions are unweighted, since we do not have person weights in our internal ACS file.) The industry sector distribution provides information on which sectors have the most employment, based on our ACS subsample, and which sectors have relatively more employment in ACS compared to LEHD data and vice versa. To the extent that a sector has more employment in ACS than LEHD data, for example, the "extra" ACS jobs would be in a different LEHD sector and would be mismatched. (Of course, there would likely be some mismatches even if the percent employment in the ACS and LEHD sectors were the same.) Overall, the industry distributions from ACS and LEHD data are similar. The manufacturing and retail trade sectors have relatively more employment in the ACS than LEHD data. Wholesale trade, management of companies and enterprises, administrative and support and waste management and remediation services, and public administration have relatively more employment in LEHD than ACS data. In the LEHD data, some multi unit employers have a headquarters unit assigned a management of companies and enterprises industry code. At those establishments, all individuals would get the management industry code, even though individuals at headquarters may understand that

---

<sup>38</sup> ACS data is coded to 4-digit CIC while QCEW data is coded to 6-digit NAICS, so the codes would not be identical for the same text, but we would expect the 6-digit NAICS to correspond with the 4-digit CIC after taking into account appropriate mapping of the codes.

<sup>39</sup> Assignment is an imputation method in which values for a missing or inconsistent item can be derived from the person's other responses to the survey. Items that cannot be derived from other responses may be imputed based on answers from other members of the household or from people believed to have similar characteristics.

they work in manufacturing, for example. We would therefore expect LEHD data to have more individuals in the management sector than ACS data. Tables 6 and 7 display the percent of main jobs in each LEHD (ACS) sector for each ACS (LEHD) sector, which provide further information on where industry mismatches occur. These tables are discussed in more detail below.

Figure 4 charts the percent of jobs in each ACS (LEHD) sector that are in the same sector in the LEHD (ACS) data. From Table 3, recall there is an overall sector match rate of 75 percent. The match rate varies across sectors and by whether sector is defined using ACS industry information or LEHD industry information. The management of companies and enterprises sector has the lowest match rate, which is not surprising, given that few ACS respondents report working in this industry or they provide insufficient information for coders to classify them appropriately (e.g., no mention of headquarters) (see Figure 3). For jobs in this industry in the LEHD data, the match rate is 2 percent. Wholesale trade has low match rates, especially using the LEHD-defined sector. The agriculture, other services, and arts, entertainment, and recreation sectors also have relatively low match rates. Finance and insurance, educational services, health care and social assistance, and accommodation and food services all have relatively high match rates, greater than 80 percent based on both ACS and LEHD sectors. Table 8 presents a summary of results by industry sector. It includes results from Figures 2 and 3 and includes a selection of results from the other industry analyses.

As discussed above, there are several potential sources of mismatch between the ACS and LEHD industry data. To examine the extent to which multi units may contribute to mismatch through our use of the SEIN employment mode industry, Table 5 presents the percent of SEINs that are multi unit by ACS and LEHD industry sector and industry sector match rates by single/multi unit and ACS and LEHD sector. Recall, 58 percent of employers in our analysis sample are single unit; 42 percent are multi unit. Of multi units, 61 percent have 4-digit CIC industry variation across units, which means that of all employers, 25 percent are multi units with industry variation across units. Of multi units with CIC variation across units, an average of 76 percent of employment is at the modal NAICS.

Overall, the sector match rate is very similar between single units and multi units, and the match rate is actually slightly higher for multi units. From Table 3, single units have an overall match rate of 74 percent and multi units have a match rate of 77 percent. From Table 4, the utilities, retail trade, finance and insurance, management of companies and enterprises, and public administration sectors all have more than half of employers as multi units.<sup>40</sup> A high percentage of multi-unit employers does not necessarily correspond with a low match rate. For ACS sectors, the correlation between percent multi unit and percent sector match is 0.20; for LEHD sectors, the correlation is 0.02. In some ACS sectors, such as construction, single units are more likely to match the LEHD industry sector, while in other ACS sectors, such as agriculture, forestry, fishing, and hunting, multi units are more likely to match the LEHD industry sector.

Table 6 shows the percent of jobs in the LEHD sectors for each ACS sector.<sup>41</sup> The diagonal, highlighted in yellow, is the industry sector match rate for the row sector. For example, for ACS sector 42 (wholesale trade), the table shows that 57 percent of jobs in sector 42 are also in sector 42 in LEHD data. Eleven percent of jobs are in LEHD sector 31-33 (manufacturing) and 12 percent of jobs are in LEHD sector 44-45 (retail trade). This suggests that it may be difficult for employees or industry coders to know whether the job is manufacturing, wholesale trade, or retail trade. It may also be difficult to code responses to these sectors in the QCEW data. Each row sums to 100. Table 7 is similar to Table 6 but shows the percent of jobs in the ACS sectors for each LEHD sector. In Table 7, the LEHD wholesale trade sector has 23 percent of jobs in the ACS manufacturing sector and 17 percent of jobs in the ACS retail trade sector, similar to Table 6.

Appendix Table 1 shows the percent of jobs in the LEHD sectors for each ACS 4-digit CIC. The first column lists the 4-digit CIC. The second column is a description of the 4-digit CIC. The third column is the percent of PIKs (equivalent to the percent of main jobs) that have that ACS 4-digit CIC. The remaining columns show the percentage of jobs in each LEHD sector; each row sums to 100. The yellow and orange cells are the match rates: the percent of each 4-digit CIC that is in the same sector in the LEHD data. Orange cells have match rates below 50

<sup>40</sup> For ACS or LEHD sectors.

<sup>41</sup> Noted earlier, Table 1 lists the twenty 2007 NAICS sectors: the numeric codes and a description of each sector. This table is useful for Tables 6, 7, and 8, which use numeric codes for the sectors, since there is insufficient space to list descriptions of each sector.

percent. Green cells show (mismatched) LEHD industry sectors that have more than 10 percent of jobs for that 4-digit CIC. Cells based on very small numbers of PIKs may be less reliable than cells based on more observations.

## 5. Missing ACS Industry Analyses

We do some additional analyses on characteristics and industry distributions for individuals missing values for the unedited industry variable (UIND). Table 9 compares the characteristics of individuals with missing industry information (UIND is blank) to those with non-missing industry information. Industry is not missing at random. Those with missing industry are more likely to be young, lower educated, lower income, Black or Hispanic, unemployed or not in the labor force, less likely to have worked in the past year, and worked fewer hours or no hours in the past year.<sup>42</sup> Many of the demographic characteristics listed in the table are used in the ACS industry imputation model to account for differences between those with missing industry and those with non-missing industry. This table is based on an ACS sample of individuals with industry in universe, where industry is edited and imputed if missing. In the sample, age is greater than or equal to 16 and the individual worked in the last five years. The ACS sample is not linked to the ACS PIK crosswalk or to LEHD data sets, and the calculations are unweighted.

Figure 5 compares the assigned or imputed industry distribution for individuals missing values for the unedited industry variable (UIND) to the industry distribution for those with non-missing values for UIND. (Industry may still be edited or imputed if UIND is missing.) Figure 5 includes both the ACS and LEHD industry sector distributions. From Figure 5, the assigned/imputed ACS industry sector distribution for those missing UIND differs from the ACS industry distribution for those with reported UIND. This is not surprising, given their different demographic characteristics.<sup>43</sup> The LEHD distribution for individuals missing UIND is similar to the assigned/imputed ACS distribution for those missing UIND. This suggests that the ACS imputation model is assigning a similar industry distribution. However, for any given individual, the industry sector match rate for those with imputed IND is very low, 14 percent (Table 5). Therefore, while aggregate industry analyses with imputed ACS data may reflect a similar distribution to LEHD data, multivariate analyses (e.g., industry by sex or earnings) may differ. The assigned/imputed ACS and LEHD sector shares are more similar in some sectors than others. For example, in the construction sector, the ACS imputation models assigns/imputes more individuals to that sector than appear in the LEHD data. In wholesale trade, the ACS assigned/imputed industry sector and the LEHD sector have different shares of the distribution, but the ACS and LEHD data have low match rates for wholesale trade.

## 6. Conclusion and Discussion

We compare person-reported industry in the ACS to employer-reported industry in the QCEW data. ACS individuals are linked to LEHD individuals using Protected Identification Keys (PIKs). The date of the ACS interview and LEHD earnings information are used to select a job in the LEHD data that likely corresponds to the job referenced in the ACS. We find an overall industry sector match rate of 75 percent and a 4-digit CIC match rate of 61 percent. Industry match rates vary by sector and by whether industry sector is defined using ACS or LEHD industry information.

The educational services and health care and social assistance sectors have some of the highest industry sector match rates. Wholesale trade has relatively low match rates, especially for the LEHD sector. Jobs in this sector in the LEHD data are often categorized in the manufacturing or retail trade sectors in the ACS data, and vice versa. Descriptions of industries in wholesale trade may be difficult to distinguish from manufacturing or retail trade. The ACS survey directly asks respondents whether their employer is manufacturing, wholesale trade, retail trade, or something else, but in the QCEW firms can review assigned NAICS codes and they verify NAICS codes in the Annual Refiling Survey. The management of companies and enterprises sector is a relatively small sector, but it has very low industry sector match rates. Analysts who want everyone employed at headquarters included in the management of companies sector, may want to use LEHD public-use statistics or microdata. Analysts who are

---

<sup>42</sup> Bollinger and Hirsch (2013) conclude that high-earning men are more likely to have earnings nonresponse in the CPS and that earnings nonresponse is not ignorable. Bollinger and Hirsch (2006) also show that the inclusion of imputed earnings in regression models leads to match bias. Earnings imputation rates in the CPS are much higher (around 20 percent in the March Annual Social and Economic Supplement) than ACS industry imputation rates.

<sup>43</sup> UIND is missing for three percent of the sample.

interested in the employer's primary activity, as reflected in the ACS, may want to use ACS public-use statistics or microdata.

Industry matching provides some information on the quality of ACS imputed industry data. For individuals with assigned or imputed ACS industry, the ACS industry sector distribution is similar to the LEHD industry sector distribution. This suggests the ACS imputation model is approximating the LEHD industry distribution. However, for individuals with imputed industry, the industry sector match rate is only 14 percent, even with an overall industry sector match rate of 75 percent (Table 5). This suggests that ACS microdata users should cautiously use imputed industry values. Public-use statistics that tabulate other variables by industry, for example, earnings by industry, may be slightly inaccurate to the extent that industry is imputed and the imputation is incorrect. Industry was imputed for only three percent of respondents in 2009, but missing response has increased since then due to a scaling back of Census Bureau survey follow-up operations for budgetary reasons. In addition, administrative data may be useful for providing information on industry when industry is not reported, with the acknowledgement that the administrative data industry distribution differs from an underlying survey industry distribution. Nonetheless, incorporating administrative data regularly into ACS data would require a substantial investment of staff time, and the project may be too costly relative to the three percent of respondents with imputed industry.

It is difficult to assess the overall accuracy of the ACS data compared with the LEHD data. While we would expect a firm to provide better information about industry than a worker at the firm (or a family member of the worker), the open-ended nature of the industry questions in the QCEW and the ACS mean that the employer or the worker may provide insufficient information to accurately code industry. In addition, the ACS and QCEW may use different procedures to assign SIC or NAICS codes from the text. To the authors' knowledge, there has been no systematic study comparing the industry coding and editing procedures of the ACS (respondent-reported) to the industry coding and editing procedures of the QCEW, or other employer-reported data. Future studies on survey-based and firm-based industry data should consider a detailed examination of data collection, coding, and editing procedures to provide additional information on potential sources of error and data mismatching, and potential advantages for using each type of data.

## **Appendix: ACS Survey Weight and Class of Worker Restriction Analyses**

In Appendix Figure 1, we examine how using ACS survey weights (person weight - PWGT) would change our industry sector distribution. As noted earlier, the pre-swapped ACS data we use does not contain person weights and we do not use them in our analyses. Figure 6 is based on data where variable values have been swapped for confidentiality protection and person weights have been calculated. We select a sample based on the same ACS variable restrictions used for our analyses (ESR=1, etc.), but it is not limited to individuals with non-missing PIKs and those that link to LEHD data. Figure 4 shows industry sector distributions using the edited and imputed IND variable. The UIND variable, without edits or imputations, produces very similar results. Using person weights increases the number of jobs in construction, retail, administrative, and support and waste management, and accommodation and food services. The use of person weights decreases the number of jobs in manufacturing, education, and health care and social assistance. Even with these small differences, the weighted and unweighted distributions are similar, and the fact that our analyses are unweighted does not appear to be substantially influencing our results. Appendix Table 2 presents the industry distributions found in Figure 5 (weighted and unweighted) and separately for the unedited and edited industry variables (UIND and IND).

The ACS publishes data on all class of worker categories. For our industry comparison, we limit the sample to private wage and salary workers and state and local government workers, excluding federal government workers, the self-employed, and unpaid family workers. In Appendix Table 3, we compare the characteristics of an ACS sample using our sample restrictions to an ACS sample of federal government, self-employed, and unpaid family workers. This sheds light on how our sample restrictions affect characteristics of the group. Both samples in Appendix Table 3 are based on swapped 2009 ACS data that are not linked to PIKs or LEHD data. Both samples are also limited to the employed, at work population (ESR=1) and individuals working in states other than Massachusetts (POWS not equal to 025). The workers in the analysis sample are younger, less educated, and less likely to have served in the military than the excluded workers. The analysis sample workers are less likely to be in management or professional occupations and construction occupations and more likely to be in production, transportation, and material moving occupations and sales and office occupations. Not surprisingly, the analysis sample workers are less likely to report self-employment income than the excluded workers, which include the self-employed.

## References

- Abowd, John and Martha Stinson. Forthcoming. "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data." *Review of Economics and Statistics*.
- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. 2005. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." LEHD Technical Paper TP-2006-01.
- Abraham, Katharine G., John C. Haltiwanger, Kristin Sandusky, James Spletzer. Forthcoming. "Exploring Differences in Employment between Household and Establishment Data." *Journal of Labor Economics*.
- Becker, Randy, Joel Elvery, Lucia Foster, C.J. Krizan, Sang Nguyen, and David Talan. "A Comparison of the Business Registers Used by the Bureau of Labor Statistics and the Bureau of the Census." Presented at the 2005 American Statistical Association Joint Statistical Meetings. <http://www.bls.gov/osmr/pdf/st050270.pdf>
- Bollinger, Christopher R. and Barry T. Hirsch. 2006. "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching." *Journal of Labor Economics*, 24: 483-519.
- Bollinger, Christopher R. and Barry T. Hirsch. 2013. "Is Earnings Response Ignorable?" *The Review of Economics and Statistics*, 95(1): 407-416.
- Bound, John and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9(1): 1-24.
- Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers. 1994. "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, 11(3): 345-368.
- Elvery, Joel, Lucia Foster, C.J. Krizan, and David Talan. "Preliminary Micro Data Results from the Business List Comparison Project." Presented at the 2006 American Statistical Association Joint Statistical Meetings.
- Fairman, Kristin, Lucia Foster, C.J. Krizan, and Ian Rucker. "An Analysis of Key Differences in Micro Data: Results from the Business List Comparison Project." Presented at the 2008 American Statistical Association Joint Statistical Meetings.
- Polivka, Anne E. and Jennifer M. Rothgeb. 1993. "Overhauling the Current Population Survey: Redesigning the Questionnaire." *Monthly Labor Review*, 116(9) (September), pp. 10-28.
- Roemer, Marc. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation." LEHD Technical Paper TP-2002-22.
- Stevens, David W. 2002. "Employment That is Not Covered by State Unemployment Insurance Laws." LEHD Technical Paper TP-2002-16.
- Stinson, Martha, Graton Gathright, and Jeremy Skog. 2012. "Comparing Job Characteristics from the 2010 SIPP-EHC Field Test to the Census Bureau Business Register." Working Paper.
- U.S. Census Bureau. 2009. *American Community Survey, Puerto Rico Community Survey, 2009 Subject Definitions*. [http://www.census.gov/acs/www/Downloads/data\\_documentation/SubjectDefinitions/2009\\_ACSSubjectDefinitions.pdf](http://www.census.gov/acs/www/Downloads/data_documentation/SubjectDefinitions/2009_ACSSubjectDefinitions.pdf)
- U.S. Census Bureau. 2009. *Design and Methodology: American Community Survey*. [http://www.census.gov/acs/www/Downloads/survey\\_methodology/acs\\_design\\_methodology.pdf](http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf)

**Table 1: 2007 NAICS Sectors**

<b>Sector</b>	<b>Description</b>
11	Agriculture, Forestry, Fishing and Hunting
21	Mining, Quarrying, and Oil and Gas Extraction
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
92	Public Administration



**Table 2: Sample Selection and Job Match Observation Counts**

	N	ACS-LEHD Match Rate	Percent of Individuals/ Jobs
(1) 2009 ACS Population File	4,534,632	--	--
<b>Populations Excluded from Analysis Sample</b>			
NOT employed, at work (ESR!=1)	2,532,851	--	--
Federal government employee (COW=5)	84,063	--	--
Self-employed not incorporated (COW=6)	190,700	--	--
Self-employed incorporated (COW=7)	94,188	--	--
Unpaid family workers (COW=8)	7,934	--	--
Work in Massachusetts (POWS=025)	43,996	--	--
(2) ACS Subsample	1,692,601	--	--
(3) ACS Subsample with PIKs	1,550,845	--	--
(4) Sample (3) that merged to LEHD PHF_B (all available years) <sup>(A)</sup>	1,531,355	99	--
(5) Sample (4) with positive LEHD earnings in the ACS interview quarter or the previous quarter <sup>(B)</sup>	1,432,008	92	--
(6) Sample (5) with at least one plausible job	1,410,181	--	--
<b>ACS Respondents in Sample (6): Percents Below Based on Line (6)</b>			
(8) With one plausible job	1,233,903	--	87
(9) With multiple plausible jobs	176,278	--	13

Note: The 2009 ACS PIK crosswalk contains some ACS observations with identical PIKs; PIKs are unduplicated before merging to the ACS population file. Populations excluded from the analysis sample are not mutually exclusive. The PIK rate of the ACS subsample is 92 percent. Plausible jobs are defined as follows. For ACS respondents with RDATE in the last two months of the quarter, LEHD jobs with positive earnings in the ACS interview quarter are considered plausible jobs. For ACS respondents with RDATE in the first month of the quarter, LEHD jobs with positive earnings in the ACS interview quarter or the previous quarter are considered plausible jobs. Therefore, individuals will be considered to have zero plausible jobs if RDATE is in the last two months of the quarter and they have LEHD earnings only in the quarter prior to the ACS interview. The main job is the plausible job with the highest earnings. See text for additional details. Percents in (8) and (9) sum to 100.

Footnotes: (A) ACS-LEHD Match Rate=100\*((5)/(4)); (B) ACS-LEHD Match Rate=100\*((6)/(4)).

**Table 3: Geography Match Rates for Main Jobs and Alternative Plausible Jobs**

	Percent Match		
	All Main Jobs	Main Jobs for PIKs with >1 Plausible Job	Alternative Plausible Jobs
State	98	96	89
County	73	68	51
County, LEHD State=Minnesota	80	73	51
Tract	47	39	13
Tract, LEHD State=Minnesota	64	56	18

Note: In the LEHD data, state place of work is at the SEIN level. County and tract comparisons are based on the first implicate of the SEINUNIT impute for multi units. In Minnesota, SEINUNIT is reported and not imputed in most cases. The number of observations varies by geographic level and job type. For the national sample, one percent of jobs in each group are missing ACS county information. 21 of main jobs are missing ACS tract information; 23 percent of main jobs with more than one plausible job and 23 percent of alternative plausible jobs are missing tract information. The LEHD program imputes county and tract for firms missing information.

**Table 4: Industry Match Rates for Main Jobs**

ACS Edited Industry Variable (IND)/		Percent Match	
LEHD SEIN Employment Mode Industry	Percent Main Jobs	Sector	4-digit CIC
All PIKs	100	75	61
<i><b>By IND Reporting Status</b></i>			
IND As Reported	97	77	62
IND Assigned	<1	65	35
IND Imputed	3	14	5
<i><b>By LEHD Single/Multi Unit SEIN</b></i>			
LEHD Single Unit SEIN	58	74	60
LEHD Multi Unit SEIN	42	77	61
ACS Edited Industry Variable (IND)/		Percent Match	
LEHD SEINUNIT Industry	Percent Main Jobs	Sector	4-digit CIC
All SEIN	100	73	58
<i><b>By LEHD Single/Multi Unit SEIN</b></i>			
LEHD Single Unit SEIN	58	73	59
LEHD Multi Unit SEIN	42	73	57

Note: For main jobs, there is one job (SEIN) per person (PIK). IND is non-missing for our analysis sample. SEINUNIT industry is based on the SEINUNIT from the first implicate of the SEINUNIT impute.

**Table 5: Industry Sector Match Rates by Single/Multi Unit**

	Percent Multi Unit		Multi Units		Percent Match					
	LEHD		Average Percent Employment at		ACS Sector			LEHD Sector		
	ACS Sector	Sector	Employment Mode Sector		All Firms	Single	Multi Units	All Firms	Single	Multi Units
	ACS Sector	Sector	ACS Sector	LEHD Sector	All Firms	Units	Multi Units	All Firms	Units	Multi Units
Agriculture, Forestry, Fishing, and Hunting	24	14	82	80	55	63	27	65	67	54
Mining, Quarrying, and Oil and Gas Extraction	42	45	81	82	67	66	68	65	69	61
Utilities	58	68	82	89	58	42	69	87	83	90
Construction	18	13	78	83	68	74	40	76	78	61
Manufacturing	37	37	81	82	73	74	72	84	85	83
Wholesale Trade	36	30	83	84	57	61	49	38	38	40
Retail Trade	61	65	87	88	78	63	87	83	75	87
Transportation and Warehousing	43	44	91	92	71	69	75	76	75	78
Information	48	48	87	91	65	57	72	67	60	74
Finance and Insurance	59	59	89	91	83	78	87	89	85	92
Real Estate and Rental and Leasing	33	31	85	89	63	62	64	62	60	66
Professional, Scientific, and Technical Services	26	23	85	84	68	72	55	67	69	61
Management of Companies and Enterprises	52	69	76	67	21	21	21	2	3	1
Administrative and Support and Waste Management and Remediation Services	34	33	89	91	62	62	61	43	43	43
Educational Services	45	48	92	88	90	89	92	87	91	82
Health Care and Social Assistance	41	42	84	84	84	85	82	87	89	85
Arts, Entertainment, and Recreation	24	25	78	68	54	57	44	68	73	54
Accommodation and Food Services	40	42	95	93	83	82	86	82	82	82
Other Services	26	18	85	90	55	61	39	58	57	59
Public Administration	57	49	67	70	77	84	71	65	62	69

Note: The ACS and LEHD sector match rates (all firms) are also displayed in Figure 4. The average percent employment at employment mode industry sector includes multi units with no variation in industry across units, which would have 100 percent of employment at employment mode industry sector.

**Table 6: Industry Mismatches: Percent Jobs in LEHD Sectors, for Each ACS Sector  
(Each Row Sums to 100)**

ACS IND Sector	LEHD SEIN Employment Mode Sector																			
	11	21	22	23	31-33	42	44-45	48-49	51	52	53	54	55	56	61	62	71	72	81	92
<b>11</b>	55	<1	<1	2	11	10	4	2	<1	<1	1	1	1	3	2	1	1	1	2	3
<b>21</b>	<1	67	1	5	9	3	1	3	<1	<1	1	3	2	2	<1	1	<1	<1	1	<1
<b>22</b>	<1	1	58	4	2	2	1	1	<1	<1	<1	2	6	4	<1	<1	1	<1	1	16
<b>23</b>	<1	1	<1	68	4	2	2	1	<1	<1	1	3	1	4	1	1	<1	1	1	7
<b>31-33</b>	<1	1	<1	2	73	8	2	1	1	<1	<1	4	2	3	<1	1	<1	1	1	<1
<b>42</b>	1	1	<1	2	11	57	12	3	1	<1	1	3	2	3	<1	1	<1	1	1	<1
<b>44-45</b>	<1	<1	<1	1	2	7	78	1	1	<1	1	1	1	2	1	2	<1	2	1	<1
<b>48-49</b>	<1	<1	<1	1	3	4	3	71	<1	<1	1	2	1	5	2	1	<1	1	1	3
<b>51</b>	<1	<1	<1	3	3	3	3	<1	65	1	1	8	1	5	1	1	1	1	1	4
<b>52</b>	<1	<1	<1	<1	<1	1	1	<1	1	83	1	3	3	3	1	2	<1	<1	1	1
<b>53</b>	<1	<1	<1	4	1	1	2	1	1	2	63	3	1	5	1	5	1	2	3	3
<b>54</b>	<1	<1	<1	1	5	4	2	1	5	2	1	68	1	5	2	2	<1	1	1	1
<b>55</b>	<1	2	1	1	12	7	16	2	2	5	3	7	21	8	<1	3	1	6	2	<1
<b>56</b>	<1	<1	<1	3	3	3	2	2	1	2	1	5	1	62	1	4	1	3	2	3
<b>61</b>	<1	<1	<1	<1	<1	<1	1	<1	<1	<1	<1	1	<1	1	90	3	<1	1	1	1
<b>62</b>	<1	<1	<1	<1	<1	<1	1	<1	<1	1	<1	1	1	2	3	84	<1	1	2	3
<b>71</b>	<1	<1	<1	1	1	1	2	<1	1	<1	1	1	<1	2	3	2	54	11	6	11
<b>72</b>	<1	<1	<1	1	1	1	4	<1	<1	<1	1	1	<1	2	1	1	2	83	1	<1
<b>81</b>	<1	<1	<1	2	3	4	9	1	1	1	1	3	1	3	4	9	1	2	55	1
<b>92</b>	<1	<1	<1	1	<1	<1	1	1	<1	1	<1	1	<1	2	8	6	1	<1	1	77

Note: The diagonal, highlighted in yellow, is the industry sector match rate for the row sector.

**Table 7: Industry Mismatches: Percent Jobs in ACS Sectors, for Each LEHD Sector**  
**(Each Row Sums to 100)**

LEHD SEIN Employment Mode Sector	ACS IND Sector																			
	11	21	22	23	31-33	42	44-45	48-49	51	52	53	54	55	56	61	62	71	72	81	92
11	65	<1	<1	2	7	7	5	2	<1	<1	1	2	<1	2	1	1	<1	2	1	1
21	<1	65	2	6	11	5	2	3	<1	<1	<1	2	<1	1	<1	<1	<1	<1	1	<1
22	<1	1	87	3	2	1	1	1	<1	<1	<1	1	<1	1	<1	<1	<1	<1	<1	2
23	<1	1	1	76	5	1	3	1	2	<1	1	2	<1	2	<1	1	<1	1	1	1
31-33	1	<1	<1	2	84	3	2	1	1	<1	<1	3	<1	1	<1	1	<1	1	1	<1
42	2	<1	<1	2	23	38	17	3	1	1	<1	5	<1	2	1	1	<1	1	2	<1
44-45	<1	<1	<1	1	2	3	83	1	1	<1	<1	1	<1	1	1	1	<1	2	2	<1
48-49	<1	1	<1	2	4	3	4	76	<1	<1	<1	2	<1	2	1	1	<1	1	1	1
51	<1	<1	<1	1	4	1	5	<1	67	2	<1	12	<1	2	1	1	1	<1	1	<1
52	<1	<1	<1	<1	1	<1	1	<1	<1	89	1	2	<1	1	1	2	<1	<1	<1	<1
53	<1	<1	<1	5	2	2	5	3	1	2	62	2	<1	3	1	4	1	3	2	1
54	<1	<1	<1	2	8	1	3	1	3	3	1	67	<1	3	1	3	<1	1	1	1
55	<1	1	6	3	28	6	11	3	3	13	2	5	2	2	1	8	1	3	2	<1
56	<1	<1	1	4	9	2	5	4	3	4	2	7	<1	43	2	7	1	3	2	2
61	<1	<1	<1	<1	<1	<1	1	1	<1	<1	<1	1	0	<1	87	4	1	1	1	3
62	<1	<1	<1	<1	1	<1	1	<1	<1	1	1	1	<1	1	2	87	<1	1	2	2
71	<1	<1	1	1	1	<1	3	1	2	<1	1	1	<1	2	3	3	68	7	2	4
72	<1	<1	<1	1	1	<1	3	<1	<1	<1	<1	1	<1	1	2	2	4	82	1	<1
81	<1	<1	<1	2	4	1	4	1	1	1	2	2	<1	2	3	10	4	2	58	2
92	<1	<1	4	6	<1	<1	1	2	2	1	1	1	<1	2	2	8	4	<1	1	65

Note: The diagonal, highlighted in yellow, is the industry sector match rate for the row sector.

**Table 8: Main Results for Each Industry Sector**

<b>Industry Sector</b>	<b>ACS, LEHD Percent of Main Jobs</b>	<b>ACS, LEHD Sector Match Rates</b>	<b>Other Main Results and Notes</b>
Agriculture, Forestry, Fishing and Hunting	1, 1	55, 65	One of the smallest sectors. For ACS sector, 11 percent jobs in manufacturing sector and 10 percent jobs wholesale trade sector in LEHD data. Comprised of six 4-digit CIC codes. Several ACS 4-digit CICs in this sector have a substantial share of jobs in the LEHD manufacturing, wholesale trade, or public administration sectors, suggesting there is insufficient information in the text descriptions or different approaches to coding these industries.
Mining, Quarrying, and Oil and Gas Extraction	1, 1	67, 65	One of the smallest sectors. This sector has a relatively low place of work tract match rate. Comprised of five 4-digit CIC codes.
Utilities	1, 1	58, 87	One of the smallest sectors. This sector has substantially different match rates depending on whether sector is defined using ACS or LEHD information. For ACS sector, 16 percent of jobs are in public administration sector in LEHD data.
Construction	5, 4	68, 76	Consists of one 4-digit CIC code.
Manufacturing	13, 11	73, 84	One of the largest sectors. This sector has more jobs in the ACS data than LEHD data and relatively high match rates. We would expect a higher match rate for the LEHD sector, given the greater reporting of this sector in the ACS data. Comprised of numerous 4-digit CIC codes. This sector has a relatively high place of work tract match rate.
Wholesale Trade	3, 5	57, 38	This sector has more jobs in LEHD data than ACS data. Much lower match rates than manufacturing or retail trade, especially in LEHD data. More than 10 percent of ACS wholesale trade jobs are in the LEHD manufacturing and retail trade sectors. For the LEHD wholesale trade sector, 23 percent of jobs are in the ACS manufacturing sector and 17 percent of jobs are in the retail trade sector.
Retail Trade	12, 11	78, 83	One of the largest sectors. Relatively low place of work tract and county match rates.
Transportation and Warehousing	3, 3	71, 76	For ACS and LEHD sectors, mismatched jobs spread over several sectors.
Information	3, 2	65, 67	For LEHD (ACS) sector, 12 (8) percent of jobs in professional, scientific, and technical services sector.
Finance and Insurance	6, 5	83, 89	Relatively high sector match rates.

**Table 8 continued: Main Results for Each Industry Sector**

<b>Industry Sector</b>	<b>ACS, LEHD Percent of Main Jobs</b>	<b>ACS, LEHD Sector Match Rates</b>	<b>Other Main Results and Notes</b>
Real Estate and Rental and Leasing	1, 1	63, 62	For ACS and LEHD sectors, mismatched jobs spread over many sectors.
Professional, Scientific, and Technical Services	6, 6	68, 67	Relatively high place of work tract and county match rates.
Management of Companies and Enterprises	<1, 1	21, 2	Lowest ACS and LEHD sector match rate of all sectors. ACS reports very few jobs in this sector. (See text for further discussion.) This sector consists of only one 4-digit CIC and therefore has identical sector and 4-digit CIC match rates. More than 10 percent of ACS management jobs are in LEHD manufacturing and retail trade sectors.
Administrative and Support and Waste Management and Remediation Services	3, 4	62, 43	This sector has more jobs in LEHD data than ACS data and has a higher match rate in the ACS data. For the ACS and LEHD sectors, the non-matched jobs are spread over many different sectors. This sector includes temporary help agencies and has relatively low county and tract place of work match rates.
Educational Services	11, 12	90, 87	This sector is one of the largest sectors and has very high sector match rates. This sector has four 4-digit CIC codes and has very similar sector and 4-digit CIC match rates. This sector has relatively low place of work tract match rates. This is likely due to underreporting of establishments (schools) in LEHD data; information is often reported at the firm (SEIN) level.
Health Care and Social Assistance	15, 14	84, 87	This sector is one of the largest sectors and has very high sector match rates. Comprised of numerous 4-digit CIC codes. This sector has some of the highest place of work tract and county match rates.
Arts, Entertainment, and Recreation	2, 2	54, 68	Higher sector match rate based on LEHD sector. More than 10 percent of ACS arts, entertainment and recreation jobs are in the LEHD accomodation and food services and public administration sectors.
Accommodation and Food Services	6, 6	83, 82	Relatively high sector match rates.
Other Services	3, 3	55, 58	Relatively low sector match rates.
Public Administration	5, 5	77, 65	This sector has more jobs in LEHD data than ACS data and has a higher match rate in the ACS data. The ACS public administration sector has 8 percent of jobs in the LEHD educational services sector and 6 percent of jobs in the arts, entertainment, and recreation sector.

Note: ACS, LEHD percent of main jobs are the same numbers presented in Figure 3. ACS, LEHD sector match rates are the same numbers presented in Figure 4.



**Table 9: Characteristics by ACS Industry Reported/Not Reported**

	Percent Respondents	
	Industry Not Reported (UIND="")	Industry Reported (UIND not equal to "")
<b>Age (AGE)</b>		
16 to 24	26	14
25 to 34	15	18
35 to 44	15	20
45 to 54	17	23
55 to 64	15	18
65+	13	7
<b>Educational Attainment (SCHLR)</b>		
Nursery school to 12th grade, no diploma	23	11
High school graduate	33	26
Some college, but less than 1 year	8	8
1 or more years of college, no degree	15	17
Associates degree	5	8
Bachelors degree	11	19
Masters degree	4	8
Professional degree beyond a bachelors degree	1	2
Doctorate degree	1	1
<b>Class of Worker, Unedited (UCOW)</b>		
Private for-profit	66	65
Private not-for-profit	6	8
Local government	5	8
State government	3	5
Federal government	3	3
Self-employed, not inc.	10	7
Self-employed, inc.	3	4
Unpaid family workers	2	<1
Private (unspecified)	2	1
Government (unspec.)	<1	<1
Self-employed (unspec.)	<1	<1
<b>Sex (SEX)</b>		
Male	52	51
Female	48	49
<b>Employment Status Recode (ESR)</b>		
Employed, at work	46	76
Employed, with a job but not at work	1	2
Unemployed	21	7
Armed Forces, at work	<1	1
Armed Forces, not at work	--	0
Not in labor force	31	15

**Table 9 continued: Characteristics by ACS Industry Reported/Not Reported**

	Percent Respondents	
	Industry Not Reported (UIND="")	Industry Reported (UIND not equal to "")
<b>When Last Worked (WKL)</b>		
Within past 12 months	61	90
1-5 years ago	26	10
Over 5 years ago or never worked	13	<1
<b>Weeks Worked Past 12 Months, Unedited (UWKW)</b>		
50 to 52 weeks	18	35
48 to 49 weeks	5	6
40 to 47 weeks	11	15
27 to 39 weeks	12	15
14 to 26 weeks	12	12
13 weeks or less	43	16
<b>Hours worked per week, Unedited (UWKH)</b>		
0 hours	15	1
1-9 hours	7	3
10-19 hours	8	6
20-29 hours	11	9
30-39 hours	11	14
40 to 49 hours	40	51
50 or more hours	8	16
<b>Service in Armed Forces, Unedited (UMIL)</b>		
Yes, on active duty now	<1	1
Yes, on active duty in past 12 months	<1	<1
Yes, on active duty more than 12 months ago	8	8
No, training only	1	1
No, never served	90	89
Yes, on active duty (unspec.)	<1	<1
No (unspec.)	<1	0
<b>Wages/Salary Income, Unedited (UWAG)</b>		
0	13	1
1 to 9,999	35	20
10,000 to 29,999	28	28
30,000 to 49,999	13	23
50,000 to 74,999	7	16
75,000 to 99,999	2	6
100,000 to 149,999	2	5
150,000 to 199,999	<1	1
200,000 to 249,999	<1	1
250,000+	<1	1

**Table 9 continued: Characteristics by ACS Industry Reported/Not Reported**

	Percent Respondents	
	Industry Not Reported (UIND="")	Industry Reported (UIND not equal to "")
<b>Self-Employment Income, Unedited (USEM)</b>		
Loss or missing	48	19
1 to 9,999	27	32
10,000 to 29,999	14	23
30,000 to 49,999	5	10
50,000 to 74,999	2	6
75,000 to 99,999	1	3
100,000 to 149,999	1	3
150,000 to 199,999	<1	1
200,000 to 249,999	<1	1
250,000+	1	2
<b>Hispanic Origin Group (HSGP)</b>		
Not Hispanic	85	89
Hispanic	15	11
<b>Disability Recode (DIS)</b>		
With a disability	15	8
No disability	85	92
<b>In Poverty (POV)</b>		
Not in poverty	82	92
In poverty	18	8
<b>Census Region (REGION)</b>		
Northeast	20	19
Midwest	24	28
South	35	33
West	21	21
<b>Veteran/Nonveteran Status (VETSTAT)</b>		
In armed services	<1	1
Veteran	9	9
Nonveteran	91	91
<b>Race (Based on TOTRACE)</b>		
White alone	70	81
Black alone	16	9
AIAN alone	1	1
Asian alone	6	4
Some other race or two or more races	7	5

Note: Calculations based on an ACS sample of individuals with industry in universe, where industry is edited or imputed if missing. In this sample, age is greater than or equal to 16 and the individual worked in the last five years or the individual is unemployed. The ACS sample is not linked to the ACS PIK crosswalk or to LEHD data sets, and the calculations are unweighted. The number of observations varies by descriptive variable and whether industry is reported or not reported.