

**The Potential for Using Combined Survey and
Administrative Data Sources to Study Internal Labor Migration**

by

**Christopher F. Goetz
U.S. Census Bureau**

CES 17-55

September, 2017

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) 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. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see www.census.gov/ces or contact J. David Brown, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies 5K034A, 4600 Silver Hill Road, Washington, DC 20233, CES.Working.Papers@census.gov. To subscribe to the series, please click [here](#).

Abstract

This paper introduces a novel data set combining survey data from the American Community Survey (ACS) with administrative data on employment from the Longitudinal Employer-Household Dynamics program, in order to study geographic labor mobility. With its rich set of information about individuals at the time of the migration decision, large sample size, and near-comprehensive ability to detect labor mobility, the new combined ACS-LEHD data offers several advantages over the existing data sets that are typically used in the study of migration, such as the Decennial Census, Current Population Survey, and Internal Revenue Service data. An overview of how these different data sets can be employed, and examples demonstrating the usefulness of the newly proposed data set, are provided. Aggregate statistics and stylized facts are generated from the ACS-LEHD data which reveal many of the same features as the existing data sets, including the decline of aggregate mobility throughout the past decade, as well as many of the known demographic differences in migration propensity.

*Any opinions and conclusions expressed herein are those of the author 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.

1 Introduction

The vision of America as a land of opportunity has long been associated with the idea of people moving freely across the country in order to seek a better fortune. While international comparisons are difficult, the U.S. population has generally been considered to have one of the highest mobility propensities in the world, which has been viewed as a hallmark strength of the labor market (Long, 1991). This depiction belies the fact, however, that migration has been experiencing a secular decline for the past 35 years or so, as well as a precipitate decline during the Great Recession (Molloy, Smith, and Wozniak, 2011). The apparent inability or unwillingness of workers to move was of great concern following the Great Recession, as it was viewed by many as a barrier to the efficient reallocation of labor, and thus contributing to the lengthy recovery. Researchers have extensively studied the determinants of labor migration, and sought explanations for its decline.

Research on these topics has typically been conducted using surveys such as the Decennial Census and the Current Population Survey (CPS), or administrative data from income tax records provided by the Internal Revenue Service (IRS). In this paper, I demonstrate how to create a novel data set that combines survey data from the American Community Survey (ACS), with administrative data on employment outcomes from the Longitudinal Employment-Household Dynamics (LEHD) data base, both produced by the U.S. Census Bureau. With this data set, we will demonstrate how many different questions in the literature can potentially be addressed in ways that were not possible with the previously available data sources.

Combining survey data with administrative records allows researchers to leverage the

strengths of each data source, and in turn compensate for the shortcomings of the other. Survey data, with a rich set of information about individuals and households, provide the variables necessary to address a myriad of economic topics of interest. However, due to the financial and logistical constraints of conducting surveys, these data sets seldomly record the same people longitudinally, and thus deny the researcher the ability study the future outcomes in which they may be interested. The longitudinal surveys that exist, such as the Survey of Income and Program Participation (SIPP) and National Longitudinal Surveys (NLS), have relatively small sample sizes that often do not afford the econometrician sufficient statistical precision to test their hypotheses. Administrative records, on the other hand, are more narrow in focus and lack the detailed information about the economic agents. However, because these data are recorded routinely and as a matter of course for the administration of large programs or efforts, they typically have a more comprehensive scope spanning multiple years, often observing the same individuals on a repeated basis.

This type of data proves particularly beneficial in researching migration, an economic outcome that lends itself to longitudinal study. Existing surveys that do include a longitudinal component, such as the CPS, do not follow an individual after they leave their residence, thus it is not possible to determine where the individual moved to. The CPS, ACS, and Decennial Census all include retrospective measures of migration, asking a respondent whether they moved in the past year, or 5 years. While this provides a measure of migration, it does not provide any information about the respondent's situation during the time preceding and during their move. While obviously some characteristics of an individual are consistent over time, like their sex, race, and date of birth, many others could well have changed. For example, when it comes to homeownership, it is

not an obvious assumption that someone who is currently observed owning or renting did the same prior to being surveyed. Likewise, it is not possible to know from these surveys what an individual's employment status was in their prior location. The LEHD database, on the other hand, is comprised of administrative records that allow researchers to track individuals at different jobs in different locations over time, but lacks much other information about the worker and their household.

Clearly, a great potential exists for bringing these two types of data together, by combining the unique details of a survey with a means of observing subsequent labor migration via the administrative records. The goal of this paper is therefore to motivate how this newly created data set can potentially improve upon some of the shortcomings of the currently available data sources, by exploiting the strengths of both survey and administrative sources. We will also present the results from some basic measures and tests using this data set, in order to compare them to the statistics derived from existing sources.

The rest of this paper proceeds as follows. Section 2 discusses the data sets that are typically employed in the literature on internal migration, and introduces the motivation behind the creation of a combined survey/administrative database. Section 3 describes the construction of the new data set and the sources behind it. Section 4 discusses the broad topics in migration that the data have been applied to, and discusses how the new data set can help address some of the shortcomings in the existing literature, including some examples. Section 5 provides comparisons between statistics derived from my new data set with those from other sources. Section 6 concludes.

2 Data Sets Used to Study Migration

Researchers have principally used three data sets to obtain precise, nationally-representative measures of migration in the United States: the Decennial Census, and more recently its annual supplement, the ACS; the Annual Social and Economic Supplement (ASEC) of the CPS; and the IRS origin-destination data.

The decennial census has provided information on migration status since 1940, asking respondents whether they lived in the same residence 5 years prior, and if not, what state they lived in at that time. This question began asking the county of previous residence starting in 1980. The ACS, which began in a beta version in 2000 and full scale starting in 2005, provides the same information as the census but on an annual frequency. As of 2010, the ACS migration question has supplanted the one on the decennial census, as the ACS was designed to eliminate the need for the decennial “long-form”. The census and ACS also provide a way to measure long-term migration, as well as return migration, by providing the individual’s state of birth. These data provide a rich details about an individual’s demographics, employment status, income, and household conditions, but do not have a longitudinal component aside from the retrospective migration question. Decennial censuses can be linked using the confidential microdata, but the ten-year gap between observations is problematic for many research questions. As the ACS is a cross-sectional sampling of approximately 3.5 million households, repeated sampling of the same individual is rare.

The CPS microdata has contained a migration variable in its ASEC supplement since 1965, although migration rates have been published using the CPS going back to 1948. The CPS contains essentially the same information as the ACS, asking respondents where

they lived in the previous year, down to the county level. The CPS is the most detailed of the Census-based surveys with a host of information about employment, income, health insurance and employer benefits. However, the CPS has a relatively small sample size, interviewing only about 100,000 households annually compared to 3 million in the ACS, rendering it less suitable for more granular measurements and data-intensive analyses. Also, while respondents in the CPS are interviewed multiple times, those who leave their residence during the interview window are not followed to their destination, thus preventing detection of subsequent migration. Nevertheless, some studies have used the fact of premature departure from the sample as a measure of residential mobility.

The IRS origin-destination migration statistics have been released since 1975, providing measures of county-to-county and state-to-state population flows. These data are derived from individual tax returns, which provide the address of primary residence for each household. Migration is defined as a year-to-year change in the location of primary residence, weighted by the number of people in each household according to the number of exemptions claimed on the return. The published statistics provide counts and rates of gross and net migration between all county pairs, making it the most comprehensive measure of household mobility. While it is not known whether tax filers are substantially different than non-tax filers in their migration propensities, it is estimated that approximately 90% of the population is covered. Being an administrative database, however, the IRS data does not include any of the detailed characteristics available in surveys. The confidential microdata that is available for purchase allows researchers to observe the income figures on the de-identified returns, but no further information about the individual or household is available.

A small number of longitudinal surveys exist that allow the user to observe individuals

in their original location and conditions prior to migrating, as well as in their destination state. The National Longitudinal Survey of Youth 1979 (NLSY79) and Survey of Income and Program Participation (SIPP) are the two surveys that have been used the most for this purpose. While nationally representative, these surveys are limited due to their small sample sizes (under 50,000 survey units), which make detailed analyses and regression estimates less precise. Nevertheless, several studies have employed these data in the study of topics such as return migration, for example DaVanzo (1983), who found that over 25% of migrators are returning to places they had previously lived.

3 Creating a New Survey-Administrative Data Source

In this section we propose a new data set to measure migration, that combines the rich individual-level information of a household survey, with the ability to track a worker's location in an administrative database. The sample comes from the American Community Survey, the largest annual survey of the American population conducted by the Census Bureau, interviewing approximately 3.5 million households annually.¹ The ACS contains a rich set of information about individuals' jobs and occupations, income and poverty status, educational attainment, housing related circumstances, health insurance coverage, and more. However, after an ACS household is surveyed it is not surveyed again, so until now there has been little opportunity to determine anything about the workers in subsequent time periods.

It is for this purpose that we combine the ACS with the LEHD employment and earnings database, a collection of Unemployment Insurance (UI) records from 50 partici-

¹The size of the ACS was smaller from 2002-2004, as the program was still in beta status at that time, however the sample size is still considerable at around 700,000 households

pating states plus the District of Columbia, which serves to generate statistics such as the Quarterly Workforce Indicators (QWI). The data are recorded at the worker-employer (i.e. “job”) level, and report the earnings that each worker receives from a given employer on a quarterly basis. These data are linked to information on the worker’s employer establishment, sourced from the Quarterly Census of Employment and Wages (ES202) database at the Bureau of Labor Statistics. UI-covered employment represents over 95% of all private wage and salary civilian jobs and nearly all of the state and local government workforce². Groups not covered by the database include the self-employed, the unemployed, and some government workers³.

With these data we can link an ACS individual via an individual identifier to all of their LEHD-covered jobs. The crucial piece of information for the purpose of measuring job migration is in the geographic location of the employee’s workplace establishment. Although the workplace for a given individual is imputed in the case of multi-unit employers, an individual is assumed to remain at the same establishment for the length of their tenure with that firm. Additionally, the imputation is largely based on the proximity of establishments to an individual’s residence. Both these features therefore serve to bias against the observation of spurious locational changes over time.

While there are many options that a researcher may consider, a straightforward definition to identify a migrator in the data is to let an indicator variable m_{ijt} be equal to 1 when the following criteria are satisfied:

1. An ACS respondent i is observed living in geographic location j during quarter t

²For more information on the data and construction of the LEHD infrastructure, see Abowd et al. (2005).

³Workers with zero earnings in a given quarter are referred to as “non-employed”, which encompasses the unemployed as well as those who are unobserved for other reasons.

2. According to the LEHD database, individual i is no longer working in location j in quarter $t + 1$.
3. Individual i is observed beginning a new job (new employee-employer pair) at an establishment located in another geographic location $-j$, within a threshold number k quarters of t .

Within these basic parameters, one could easily vary the definition across the selection of geographical level j (region, state, metro area, etc.), time threshold k , as well as vary the sample by age group, sex, education, head of household status, or any number of stratifications. Other qualifications may be made as well to hone the definition of migration. For instance, in the case of multiple new jobs beginning at the same time, one might choose to consider only the job with the highest earnings.⁴ Also, if measuring inter-CBSA moves, a researcher might not want to count a move to a geographically adjacent CBSA that is considered to be part of the same Combined Statistical Area (CSA), which is closely linked by commuting ties. Similarly, if studying inter-state migration, the researcher may not want to count a move across state borders that fall within the same metropolitan area. The precision of the LEHD data allows for great flexibility in these sorts of definitional decisions.

Once a definition of m_{ijt} is settled on, the overall migration rate can be defined as $(\sum_{i=1}^N m_{ijt})/N$. When considering this rate, note that geographic coverage of the LEHD data varies over the sample period, as more and more states began submitting data throughout the 2000s. With 45 states included by 2002, this presents a reasonable starting point at which near-comprehensive coverage of migration can be obtained. However,

⁴This definition of a “dominant” job is commonly used within the LEHD universe.

the the number of states rose to 49 by 2004, and finally to 50 plus the Distric of Columbia starting in 2010. As a consequence, the fraction of U.S. private employment covered by the data grew from 87% in 2002 to over 95% by 2010 (Abowd and Vilhuber, 2010). Due to this fact, it is prudent to limit the sample of ACS respondents to those who are residing in a state that is LEHD-covered at the time of the ACS interview. However, the incomplete coverage clearly affects the rate at which migration is detected, thus systematically suppressing the observed migration propensities, especially during the early part of the sample ⁵.

4 Applying Theories of Migration to the Data

4.1 Prior Literature

The existing data sources described above came about at a time when theoretical models of migration were starting to take shape. Much of the early literature described the phenomenon as being driven by geographical differences in the supply and demand for labor. Seminal studies such as Lewis (1954) introduced a 2-sector model where equilibrium is restored by workers moving from the low-wage and less capital-intensive agriculture sector, to the high-wage and more capital-intensive manufacturing industry. These theories helped explain the movement from rural areas to cities that the U.S. had experienced earlier in the century, as well as spawned a new field of study in international development.

The micro-economic foundations of migration focus on the incentives of individual

⁵Any regression analyses using these data would be advised to use time dummy variables to control for the mean differences in migration propensity

workers to choose the location that will maximize their utility. Todaro (1969) and Harris and Todaro (1970) described workers as seeking to maximize expected income, while taking into account the probability of unemployment. Sjaastad (1962) discussed the costs of migration more explicitly, stating that the increase in expected lifetime earnings must exceed both the monetary and non-monetary costs of migrating.

While this basic neo-classical framework has continued to hold for decades, in the latter part of the century the focus of the literature turned to identifying more specific determinants of migration, as summarized in Greenwood (1997). Many studies have explored the mobility of different demographic groups, showing evidence that young, higher skilled, and educated workers are more likely to move (Sjaastad, 1962; Borjas, 1991; Malamud and Wozniak, 2010). Household composition has also been considered, with married couples and families with children appearing less mobile in the data. The regional differences in amenity values has also been found to influence migration, including factors such as public spending and services (Tiebout, 1956), or natural amenities like climate and geography (Knapp and Graves, 1989; Mueser and Graves, 1995). Some studies have also focused on the cost side of the decision, including the increased burden from moving longer distances (Davies et al., 2001), and the risks associated with uncertainty (Schaeffer, 1988). Other financial factors, such as housing costs and mortgage rates, have also been found to be important (Cameron and Muellerbauer, 1998; Jackman and Savouri, 1992).

Research into the forces behind migration has become particularly important in light of the dramatic drop in mobility during the Great Recession, as well as the long-term decline over past decades, leading researchers to investigate possible explanations⁶. Com-

⁶This topic is reviewed extensively in Molloy et al., (2011).

positional explanations have been sought, especially with the aging of the U.S. population and the fact that older workers tend to move less, but these types of differences across demographic groups are not of a sufficient magnitude to account for much of the overall decline (Molloy et al., 2011). The business cycle has been shown to be important with migration rates exhibiting pro-cyclical patterns, although this obviously does not explain the secular decline, nor does it go far in accounting for the precipitate decline during the Great Recession (Greenwood et al., 1997, Molloy et al., 2011).

4.2 Benefits of New ACS-LEHD Dataset in the Study of Migration

The new data set described in this paper has many advantages to offer in the study of such topics in the migration literature. First, by observing the individual in their original location prior to potentially moving, the data enable the researcher to more accurately account for the factors that enter into the migration decision. This is particularly important for addressing factors that can easily change between the origin and destination locations, such as employment and income, housing-related factors, health insurance status, family situation, and many other topics of interest. Secondly, it provides a direct measurement of labor mobility, in contrast to other sources that measure household mobility. After all, in terms of understanding the effect of migration on the labor market, economists are typically most concerned about whether workers are able to match with firms located in other areas, regardless of whether a change in residences actually occurs. Also, this data permits a more flexible definition of migration in terms of timing thresholds, especially due to the quarterly nature of the LEHD data. Instead

of being bound by the ACS or IRS definition of annual migration, the researcher can use higher-frequency measures as needed for the purposes of the given research question. Finally, the large size of the ACS sample provides high degree of precision for conducting many types of detailed analyses. This proves particularly useful for data-intensive statistical techniques, as well as allows for the econometrician to gain high levels of precision. A couple of examples that follow will highlight these benefits.

4.2.1 Example 1: Role of the Housing-Related factors

One potential cause for the decline in mobility is the historic collapse of the housing market leading up to and during the Great Recession. The phenomenon of homeowners appearing to be trapped in their homes has been well-noted since Osborne (1996). Moreover, seminal work by Genesove and Mayer (1997; 2001) showed that homeowners who had experienced a decline in their home equity were more even reluctant in selling their homes. The possible implications for long-distance migration led to an active literature in the topic, especially after the historic housing collapse of the mid-2000s. While some earlier studies provided evidence of declining home prices and “negative equity” hampering mobility, such as Chan (2001), Engelhardt (2003) and Ferreria, Gyourko, and Tracy (2008), more recent studies of the period surrounding the Great Recession and housing bust have tended to find no relationship (Dennet and Modestino, 2011; Molloy et al., 2011; Schulhofer-Wohl, 2012; Farber, 2012; Valletta, 2013).

This literature, however, suffers from a few limitations that this newly created data is well-suited to address. First of all, mobility in most of these studies is detected ex-post, and thus nothing is known about homeownership in the original location. While some studies make assumptions about prior homeownership, these analyses are flawed if

former homeowners become renters in their new location in such a way that correlates with their prior negative equity status. Because our new data set is constructed in such a way that allows us to observe an individual's migration outcome subsequent to being surveyed in their original location, we have direct information about their ex-ante homeownership and equity status. Secondly, all of the above studies depend on some kind of cross-geographic variation in housing prices, by city, state, or region. Thus they are unable to fully account for many of the differences between local labor markets, which may be correlated with local house prices and potentially bias the results. Unlike the data sets used in these prior studies, our new data set has both the geographic detail and the sample size required to exploit within-labor market variation in house prices and mobility, and thus compare owners and renters who are exposed to the same labor market at the same time. Finally, while the previous studies based on surveys or tax data measure residential mobility, the administrative jobs data that I use allow me to detect where their subsequent jobs are located, thus providing a direct measure of labor mobility. While residential and labor mobility are clearly correlated, they are not necessarily the same thing (perhaps due to factors such as long-distance commuting and telework) and thus the ACS-LEHD data set allows us to observe the outcome most relevant to the labor market.

Goetz (2013) conducts a test of the lock-in hypothesis by linking in information on homeowners' home equity, and comparing their migration propensities in the ACS-LEHD data set to comparable renters in the same location and time period. Results find that home owners with negative equity are approximately 20% less likely to relocate for jobs.

4.2.2 Example 2: Outcomes for the Unemployed

One subgroup for whom migration is posited to be particularly important is the unemployed. This topic has been a particular concern of economists in the wake of the Great Recession, who worry that the observed reluctance of the unemployed to move away from distressed labor markets could prolong their unemployment spells and thus impede the economic recovery. However, the notion that workers who choose to move will be better off than had they remained is relatively underexplored in the literature. Kennan and Walker (2011) employ a structural estimation using the NLSY79 and find that locational decisions are consistent with the maximization of lifetime income, but they do not consider the unemployed in particular, or any shorter-term incentives that workers may have for moving. Studies focusing on the duration of unemployment spells of those who migrate compared to those who stay in place, have generally found zero or negative effects of migrating (Shumway, 2000; Pekkala and Tervo 2002).

Because the new ACS-LEHD data set is able to provide information on unemployment status in the ACS, along with the subsequent employment outcomes in the LEHD data, it is ideally suited to this task. The ability to observe the unemployment status of the worker before making their migration decision is advantageous compared to using the CPS or other sources that only observe migrants after the fact. While datasets such as the Displaced Worker Survey provides a relevant set of potential job seekers, the sample is small and not necessarily representative of the unemployed as a whole. Also, because the decision to change locations is endogenous to future employment outcomes, studying such a subject typically requires the employment of selection-correction techniques. Longitudinal data sets such as the NLSY may be too small for such numerically-intensive

estimation. In fact, the only prior study that attempts to systematically control for the endogeneity problem uses a large administrative database from outside the U.S. (Pekkala and Tervo, 2002).⁷ Goetz (2014) attempts to replicate this study using the ACS-LEHD data from the U.S., and finds that unemployed workers who move for their next job indeed appear to have shorter unemployment spells than those who remain in the same area as their previous job. In particular, results suggest that those workers who had the poorest prospects of finding re-employment gained the most from relocating.

5 Comparing Measures of Migration

While the new combined ACS-LEHD microdata is clearly well suited to addressing a wide variety research questions, it is also of interest to see whether it can generate statistics that are sensible and in line with previous findings. In this section we will introduce the measures of migration that are typically used in the literature, and then compare features of the ACS-LEHD data set with those from other sources.

5.1 How migration is measured

When measuring migration, the first step is to define what sort of geographic transition constitutes a “move” of interest. Because the theoretical framework describes migration as an attempt to expose oneself to different economic conditions that are superior to one’s present situation, economists are generally concerned with measuring moves that represent a substantial change of an individual’s local labor market.

⁷Pekkala and Tervo (2002) uses a residential and jobs database maintained by the government of Sweden

Some standard definitions of labor markets have been created by government statistical agencies such as the concept of a Metropolitan Statistical Area (MSA), now known as a Core Based Statistical Area (CBSA), which groups collections of counties connected to a common urban center based on commuting patterns. Several of the publically-available data sets described above are capable of providing CBSA-level information, although changing definitions of CBSA boundaries cause some complications for researchers. With county-level information, however, CBSAs can be reconstructed even if the data source does not explicitly report them. Other narrowly defined labor market areas include Economic Areas (EA), Public-Use Microdata Areas (PUMA), and Workforce Investment Areas (WIA), although these are generally less available in public data.

Because these definitions of labor markets do not cover the entire country, by generally excluding rural areas, any statistics aggregated to these levels will by definition be less than comprehensive. As such, many aggregated statistics report cross-state migration, thus capturing all migration within US borders in a way that is consistent over time. Despite the fact that some labor markets span across state borders, state-to-state moves also generally represent a substantial change in one's employment situation and economic conditions. Tracking moves across groups of states composing the four Census Regions also generates broad and meaningful measures of national migration. These types of definitions also provide the best point of comparison between different data sources, as nearly all data sets provide information at least at the state level, whereas they may not report narrower units of geography. Because the LEHD microdata provides geographic detail about the establishment where an individual works down to the Census block, these measures can be aggregated to virtually any geographic level.

Additionally, researchers must select the timespan over which to consider migration. This selection may obviously depend on the topic at hand, although different data sets offer more or less flexibility in terms of timing. In the Decennial Census, ACS, and CPS the time period is defined as either 1 year or 5 years prior to the survey date, with no further ability for the researcher to vary the timeframe. However, those surveys that include the individual's place of birth do allow for a measure of lifetime mobility or return migration, although place of birth does not necessarily denote a location where someone has spent an extended period of time. Also note that with these retrospective questions it is possible that there have been multiple moves since the reference period, and thus some migration may go unobserved, especially in the case of the 5-year migration question. This concern is true for all measures based on two distinct start and end points, rather than a continuous longitudinal tracking.

The longitudinal data sources, on the other hand, provide more options for selecting a time span over which to detect migration. The SIPP interviews participants every 4 months, allowing for a high-frequency definition of migration, although the same individuals are tracked for only 2-5 years depending on the particular panel. The NLSY79 tracks participants for a long period of time, but interviews are only conducted every two years as of 1994. The confidential IRS microdata allows the researcher to track the location of tax filers (not dependents) annually, but the lack of other covariates limits the possible analyses. Our new data set, on the other hand, takes advantage of the quarterly frequency of the LEHD employment data, allowing for a higher-frequency measure of migration than other sources. Although designed as an annual survey, the ACS microdata also provides a specific date of interview, allowing the original time of observation to be placed in a specific month and quarter as well. Therefore, any individual appearing in

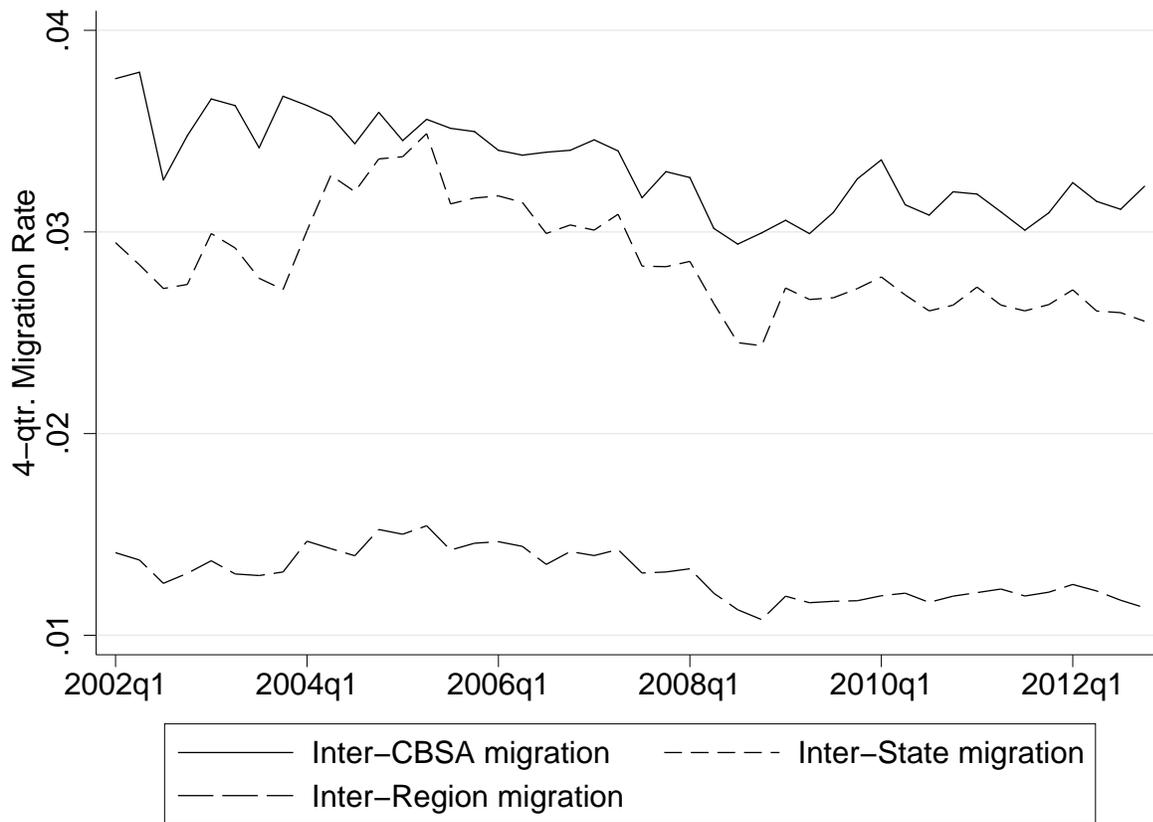
the ACS at any point can be tracked quarterly in the LEHD data for all subsequent time periods.

5.2 Results and Statistics from ACS-LEHD data

Next, we look at aggregate migration statistics calculated using some of the data sets described above. In particular, we will see how these rates have changed over time, and how the patterns and correlations from my new data source compare to those that are typically used. For reference, statistics from the other sources are taken from the work of Molloy et al. (2011), which provide a comprehensive overview of facts about migration in the United States.

The fraction of people who are observed relocating during the course of a year is small but significant, and has been found to be fairly consistent across different data sources. In order to make comparisons with existing measures, Figure 1 displays the annual migration rates as calculated from my new ACS-LEHD data set, for the time period representing 2002-2012. For this basic measure of job mobility, we will focus our sample on household heads and their spouses between the ages of 18-65. For comparability to an annual rate, we shall define migration as beginning a new job in the LEHD data in a different location than where one was originally observed in the ACS, within four quarters of the interview date. The figure below displays the average mobility rate over the sample period.

Figure 1: Migration Rates from Combined ACS-LEHD Data



Notes: Figure displays the mean national migration rate using a 4-qtr threshold. Sample consists of approximately 7.3 million ACS individuals from 2002-2012, either household head or spouse age 25-54. Individuals are weighted by the inverse proportion of their probability of appearing in the ACS.

The rate at which individuals move across broad regions, (defined as the four Census Regions of Northeast, Midwest, South, and West), has ranged between 1.1 and 1.5% and has been declining over much of the sample period. While the rate appears to rise slightly over the first part of the decade, note that this coincides with a time in which new states were being added to the LEHD data, thus mechanically increasing the detection of migration. By 2005, however, when nearly all states were included, the rate begins to nonetheless fall. The steepest period of decline occurred between 2005-2008, coinciding with the housing crisis and beginning of the Great Recession, with some stabilization in the rates occurring thereafter. These magnitudes and features of the data are similar to those found by Molloy et al. (2011), especially to those derived from the ACS and IRS data. As noted by those authors, the CPS migration measures are consistently lower than in all other sources during recent years, although the precise reasons are unknown. While the time-series of the ACS-LEHD statistics only go back to 2002, the declining rate appears to be a continuation of the decline that has been observed in the CPS going back to at least 1980, with most measures also displaying a particularly steep decline during the Great Recession.

Inter-state migration rates have hovered around 3% on an annual basis, implying that about half of inter-state moves occur within the same region of the country. Inter-CBSA moves are somewhat higher at 3-4% annually, reflecting the fact that some individuals are observed changing labor markets within the same state. These numbers are roughly comparable to the corresponding measures in the ACS and IRS data, albeit higher by a couple tenths of a percentage point.⁸ Note however, that while my sample measures the job mobility of individuals aged 18-65, other sources generally measure household

⁸For comparison, see Figure 2 in Molloy et al. (2010).

mobility weighted by the number of residents, who may be younger or older than the 18-65 age range. In other words, it should perhaps not be surprising that a migration rate based on the observed job mobility of the working age population is somewhat higher than the residential mobility rate of the population at large.

In order to make more detailed comparisons between data sources, Table 1 reports the average annual migration rate for various demographic and socio-economic subgroups in the ACS-LEHD data set compared to the corresponding calculations from the CPS during the first decade of the 2000s.⁹ While we see that the rates calculated from the ACS-LEHD data set are uniformly higher than those in the CPS, due to reasons mentioned above, the ratio between the two sets of statistics across the many subgroups is fairly consistent, with the ACS-LEHD rates ranging approximately from 50-100% higher. Moreover, the general patterns within the demographic breakdowns largely mirror one another in the two data sources.

⁹Due to the start date of the LEHD data, however, the decade is truncated by 1 year in the new data set.

Table 1: Migration Rates by Subgroups: ACS-LEHD Dataset vs. CPS

	ACS-LEHD Dataset: 2002-2010	CPS: 2001-2010 (Molloy, Smith and Wozniak)
Sex:		
Female	2.5	1.6
Male	3.4	1.7
Age:		
18-24	2.5	3.0
25-44	1.8	2.2
45-64	1.4	1.0
Race:		
White	3.2	1.8
Black	2.8	1.7
Education:		
Less than high school	2.7	1.0
High School	2.4	1.2
Some college	2.8	1.5
College+	3.5	2.1
Employment Status:		
Employed	3.0	1.6
Unemployed	5.2	3.5
Not in Labor Force	2.3	1.5
Income:		
Bottom 50%	2.8	1.7
Top 50%	3.0	1.6
Children in the household:		
No children	3.1	2.0
At least one child	2.7	1.4
Homeownership:		
Renters	4.7	3.5
Owners	2.2	0.9

Notes: The left column reports mean 4-qtr inter-state migration rates for the given subgroup across the entire period in the combined ACS-LEHD data set, weighted by the ACS survey weight. The right column reports the corresponding 1-yr (retrospective) migration rate from the CPS, as calculated by Molloy, Smith and Wozniak (2011). All stratifying variables are based on similar definitions between the two surveys.

Next we employ a simple regression model to more systematically explore which factors appear to be correlated with the propensity to move for jobs. Results from a linear probability model, which control for demographics, employment, earnings, and other factors, are presented in Table 2.

The large demographic differences in migration propensities are evident in the results. Females are found to be 0.7 percentage points less likely to migrate than males, and workers in the 45-54 age group being 0.6 percentage points less likely to move than those in the excluded 25-34 category, both consistent with the previous literature. Also, African-Americans exhibit a slightly lower migration propensity, although the gap is not nearly as large as appears in the raw averages from Table 1. Household factors are also significant, with households composed of married couples and those with children present exhibiting lower mobility. The effects of homeownership are even stronger, however, with homeowners displaying a migration propensity that is 0.4 percentage points lower than renters, corresponding with the hypothesis that homeowners are relatively fixed in place. There are also stark differences across employment statuses, with the unemployed having a migration propensity that is a full 5.5 percentage points higher than the employed – a phenomenon that is somewhat obscured in the raw averages. This is consistent with the notion that the unemployed use migration as a way to escape joblessness, as posited in the literature. Mobility also appears to be significantly lower for the highest earnings quartiles, while education level does not appear to have a significant correlation. While not all of these stylized facts are perfectly consistent with previous findings in the literature, it is reassuring that the statistics and correlations derived from the new ACS-LEHD data source are broadly in line with those generated from the previously existing data sources.

Table 2: Determinants of Migration

<i>Var</i>	$\beta/(s.e.)$
<i>Female</i>	-0.007*** (0.001)
<i>Non – white</i>	-0.001 (0.001)
<i>Married</i>	-0.001*** (0.000)
<i>Children</i>	-0.002*** (0.001)
<i>Homeowner</i>	-0.004* (0.002)
<i>Unemployed</i>	0.055*** (0.004)
<i>Not_in_labor_force</i>	0.006*** (0.001)
<i>Age35 – 44</i>	-0.003*** (0.001)
<i>Age45 – 54</i>	-0.006*** (0.000)
<i>High_school</i>	-0.001** (0.001)
<i>Some_college</i>	-0.001 (0.001)
<i>College+</i>	-0.000 (0.001)
<i>Earn_quartile_2</i>	-0.001 (0.001)
<i>Earn_quartile_3</i>	-0.007*** (0.001)
<i>Earn_quartile_4</i>	-0.006*** (0.001)
<i>N</i>	~7,300,000

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Notes: Table reports results from regression of one-quarter migration on a set of demographic variables. Sample consists of approximately 7.3 million ACS individuals from 2002-2012, either household head or spouse age 25-54. CBSA*time dummies are included in order to control for location-specific time shocks. Standard errors are clustered by CBSA.

6 Conclusion

This paper has discussed the different sources of data that are employed in the research of internal migration, and motivated the use of a novel data source that combines survey and administrative sources in order to address some shortcomings of the existing datasets. In contrast to the other data sources, this combined ACS-LEHD data simultaneously provides a flexible measurement of labor mobility, a large sample size, and the crucial ability to observe the conditions of individuals prior to the migration decision. This allows the researcher to more accurately study the determinants of migration that may change over time, such as employment or housing status. The typical data sources used in the field are not as well-suited to addressing these topics, as they either detect migration retrospectively and thus do not record these sorts of factors at the original location, or else are too small to provide a high degree of statistical precision. These data also provide a direct measurement of job-related labor mobility, rather than simply household mobility, which may be of more relevant to study of labor markets. The combined ACS-LEHD data are also shown to exhibit similar aggregate statistics and stylized facts as the existing sources. Further work is needed to thoroughly explore the source of the differences between these various data sets used to study migration, in order to properly weigh the pros of cons of using each in a given situation. Moreover, there may exist other opportunities to combine survey and administrative data, such as tax records, in ways that could achieve similar benefits as the data set proposed here. Nevertheless, the potential to employ this type of data to gain new perspectives on the causes and consequences of labor mobility in the U.S. economy is promising.

References

- [1] Cameron, G. and J. Muellbauer (1998): “The Housing Market and Regional Commuting and Migration Choices,” *Scottish Journal of Political Economy*, 45(4), 420-46.
- [2] Davies, P., Greenwood, M. and H. Li (2001): “A Conditional Logit Approach to U.S. State-to-State Migration,” *Journal of Regional Science*. 41(2), 337-360.
- [3] Ferreira, F., Gyourko, J. and J. Tracy (2010): “Housing Busts and Household Mobility,” *Journal of Urban Economics*, 68, 34-45.
- [4] Genesove, D. and C. Mayer (1997): “Equity and Time to Sale in the Real Estate Market,” *American Economic Review*, 87(3), 255-269.
- [5] Genesove, D. and C. Mayer (2001): “Loss-aversion and Seller Behavior: Evidence from the Housing Market,” *Quarterly Journal of Economics*, 116(4), 1233-60.
- [6] Goetz, C. (2013): “Falling House Prices and Labor Mobility: Testing the Lock-In Hypothesis with Matched Employer-Employee Data,” *U.S. Census Bureau, Center for Economic Studies Working Paper Series*, 13-43.
- [7] Goetz, C. (2014): “Unemployment Duration and Geographic Mobility: Do Movers Fare Better than Stayers?” *U.S. Census Bureau, Center for Economic Studies Working Paper Series*, 14-41.
- [8] Graves, P. and P. Mueser (1995): “Examining the role of economic opportunity and amenities in explaining population redistribution,” *Journal of Urban Economics*, 37(2), 176-200.
- [9] Greenwood, M. (1975): “Research on Internal Migration in the United States: A Survey,” *Journal of Economic Literature*, 13, 397-433.
- [10] Greenwood, M. (1997): “Internal Migration in Developed Countries,” in “Handbook of Population and Family Economics,” ed. M.R. Rosenzweig and O. Stark, 647-720. *Elsevier Science*, New York.
- [11] Harris, J. and M. Todaro. (1970): “Unemployment and Development: A Two-Sector Analysis,” *American Economic Review*, 60(1), 126-42.

- [12] Jackman, R. and S. Savouri (1992): “Regional migration versus regional commuting : the identification of housing and employment flows,” *Scottish Journal of Political Economy*, 39(3), 272
- [13] Knapp, T. and P. Graves (1989): “On the Role of Amenities in Models of Migration and Regional Development,” *Journal of Regional Science*, 29(1), 71-87.
- [14] Long, L. (1991): “Residential Mobility Differences Among Developed Countries,” *International Region Sciences Review*, 14(2), 133-147.
- [15] Malamud, O. and A. Wozniak (2012): “The Impact of College Education on Geographic Mobility,” *Journal of Human Resources*, 47(4), 913-950.
- [16] Molloy, R., Smith, C. and A. Wozniak (2011): “Internal Migration in the United States,” *Journal of Economic Perspectives*, 25(3), 173-196.
- [17] Molloy, R, Smith, C..and A. Wozniak (2014): “Declining Migration within the U.S.: The Role of the Labor Market,” *IZA Discussion Paper* 8149, Institute for the Study of Labor (IZA).
- [18] Schulhofer-Wohl, S. (2012): “Negative Equity does not reduce homeowners’ mobility,” *Federal Reserve Bank of Minneapolis Quarterly Review* 35(1), 2-14..
- [19] Sjaastad, L. (1962): “The Costs and Returns of Human Migration,” *Journal of Political Economy*, 70(5), 80-93.
- [20] Thiebout, C. (1956): “A Pure Theory of Local Expenditures,” *Journal of Political Economy*, 64(5), 416-424.
- [21] Todaro, M. (1969): “A Model of Labor Migration and Urban Unemployment in Less Developed Countries,” *The American Economic Review*, 59(1), 138-148.
- [22] Valetta, R. (2013): “House Lock and Structural Unemployment,” *Labour Economics*, 25(3):86-97.