New Approaches to Creating Data for Economic Geographers

Matthew Freedman[1], Julia Lane[2], and Marc Roemer[3]

Policymakers, faced with increasing demands to make decisions at a local level, are turning to statistical agencies to provide local data. Advances in matching technology, combined with the reduced cost of archiving, indexing, storing, and curating large-scale datasets, now mean that it is technically feasible to provide information at fine levels of geographic detail by means of combining administrative and survey datasets at lower cost and with potentially greater coverage. This article describes an approach that uses administrative data from U.S. unemployment insurance records to enhance the coverage and accuracy of work location information in the American Community Survey.

Key words: Administrative and survey record integration; spatial analysis; matching; imputation; commuting.

---

[1] Cornell University. Contact Information: 359 Ives Hall East, School of Industrial and Labor Relations, Cornell University, Ithaca, New York, 14850. Email: mf439@cornell.edu.

[2] National Opinion Research Center, University of Chicago. Contact Information: Suite 2000, 55 E. Monroe Street, Chicago, Illinois, 60603. Email lane-julia@norc.uchicago.edu

[3] U.S. Census Bureau. Contact Information: U.S. Census Bureau, Data Integration Division, Room 6H108E, 4600 Silver Hill Road, Suitland, Maryland, 20746. Email: marc.i.roemer@census.gov.

Disclaimer and Acknowledgment: This article reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program, which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01~AG018854-02, and the Alfred P. Sloan Foundation. The views expressed on statistical, methodological, and technical issues are those of the author(s) and not necessarily those of the U.S. Census Bureau, its program sponsors, or its data providers. Some of the data used in this article are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (http://www.ces.census.gov). For questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Attn: Holly Brown, Room 6H136C, 4600 Silver Hill Road, Suitland, MD 20746, U.S.A. (did.local.employment.dynamics@census.gov, http://lehd.did.census.gov). This article benefited from the helpful comments and suggestions of Ronald Prevost, as well as feedback from John Abowd, Waleed Almousa, Larry McGinn, George Putnam, and other LEHD Program staff.

1. Introduction

"On Tuesday, Google launched Google Earth, a free software package that gives detailed, 3-D views of cities across the globe replete with thousands of restaurants, schools, hotels and other establishments. It also provides amusing motion graphics of the route between locations or 3-D cities" (Terdiman 2005).

"…Then I turned on the overlays. Now I know how many murders were in Queens in 2000 (187), the population (2,229,379) (even if the overlay icons were on top of each other, forcing me to shut one off to see the other), see all the pharmacies, gas stations, restaurants, libraries, parking lots, hospitals, banks, airports, fire stations, parks, stadiums, golf courses, churches, trains, movie rental stores, malls, bars, and highway exits in the area, and find out where my ZIP code congressional or school district ends. Unbelievable. Amazing, absolutely amazing. You will be showing this off to your friends. I'm probably coming off as a little too excited, true, but I just genuinely enjoy using this product" (Richardson 2005).

Public demand in the United States for detailed local data seems insatiable. In the private sector, Google Earth debuted and then crashed from overwhelming public use. In the public sector, the President's 2006 budget requested $169.9 million to finance the largest-scale survey ever funded to provide local information: the American Community Survey (ACS). Pressing domestic policy matters fuel part of this demand; decisions must be made to address emergency preparedness, transportation, economic development, and workforce development issues, to name just a few. But private needs also drive part of this demand; local businesses need information about the location of potential clients and employees, and workers need to find jobs.

Providing high-quality local information on both place-of-work and place-of-residence of workers is thus a highly visible challenge for National Statistical Offices (NSOs). However, although place-of-residence information is typically of quite high quality, there are substantial problems associated with collecting place-of-work information using surveys, most notably nonresponse and misreporting. These problems, which tend to vary systematically across individual characteristics, industries, and location, could distort results at highly disaggregated geographic levels, potentially leading to incorrect inference and hence inappropriate policy decisions.

One possible approach to improving data quality is to use administrative data that are already collected on employers' addresses. In the U.S., as in most countries, employers file administrative tax information on their employees. Since such filings include each employer's address, these administrative data can be combined with survey data, and the NSO can, in principle, use the employer's actual address to check the veracity and potentially enhance the quality of the employee's response regarding place-of-work. Such an approach to data quality improvement has become more attractive for several reasons. Technological advances in collecting, indexing, archiving, and curating administrative records have greatly improved their quality and coverage. In addition, the recognition of their potential value for providing local information has led to new initiatives to improve the geographic information on the records. Finally, recent improvements in matching technologies have led to more precise and reliable integration of multiple data sources.

This article describes the approach used and lessons learned from an effort at the Longitudinal Employer Household Dynamics (LEHD) program at the U.S. Census Bureau to integrate the ACS with administrative records. We find that administrative records can be used to geocode over 40% of workplaces that would have required clerical coding, and identify as many as 80% of workplaces that would otherwise have been left blank. The approach results in identifying greater geographic dispersion of workplaces and greater geographic detail on rural employment. It particularly improves the place-of-work geocoding for workers living in poor households.

2. Background

There is substantial evidence documenting the need for local data, ranging from government agencies that need detailed data for transportation planning, land-use development, emergency preparedness, and identifying areas experiencing problems of spatial mismatch, to private sector entities requiring such data for finding workers and clients.

Transportation planning is one area in which local data are indispensable, and the need for more reliable and comprehensive information is growing. Despite many efforts to improve transportation infrastructure, congestion in America's urban areas continues to worsen, causing some 3.7 million hours of travel delay, wasting 2.3 billion gallons of gasoline, and costing the public over $63 billion in 2003 according to Lomax and Schrank (2005). They argue that allocating funds to transportation projects is not enough; policymakers must carefully target investment in order to improve mobility, which in turn requires accurate and finely detailed geographic information on where people live and work.

There is also a literature documenting the need for data to quantify imbalances between jobs and housing and related spatial mismatch issues (Pugh 1998; Horner 2004). Such imbalances arise in regions where the number of jobs is disproportionate to the local housing supply, or in regions where local residents are over- or under-qualified for the jobs available nearby. Local data can be used to characterize the location of low-income neighborhoods together with the location of fast-growing job opportunities to improve the job service efficiency of the Department of Labor. As Andersson et al. (2005) point out, data that provide localized information on hiring and wages for very detailed industries can be used by local labor market intermediaries (or staff in "One Stop" offices) who hope to place low-income workers into better jobs and sectors. Indeed, as the Deputy Governor of Illinois pointed out in a letter to the director of the U.S. Census Bureau on February 8, 2002, "Local Workforce Investment Boards (LWIBs) in Illinois, like all others in the nation, are faced with a critical need to provide their One Stop Centers with information about their workforces, their jobs, and their economic development prospects. Detailed data at the county and the LWIB level are simply not available on a timely basis, if at all."

Similar data can be used to improve the effectiveness of welfare programs. The U.S. Department of Health and Human Services used information about welfare recipients integrated with administrative data to demonstrate how a mismatch between the location of low-wage workers and jobs could affect the likelihood of individuals transitioning out of welfare dependency (Andersson et al. 2004). This report also documents a substantial and growing mismatch between worker and job location; in particular, the authors find that while job quality in most neighborhoods has improved, this is not the case in low-wage neighborhoods.

The need for local data is even more imperative given current concerns about security and emergency preparedness. Finely detailed place-of-work data are necessary to identify the location and characteristics of the population during the day, when most people are at work or in school. This information is critical to planning for emergency situations, including ensuring adequate security in high-density areas and establishing evacuation routes. Indeed, Executive Order 12906 (1994) established the National Spatial Data Infrastructure and tasked the Federal

Geographic Data Committee with developing procedures and working to implement the distribution of digital geospatial data. Such data are used by a growing number of government agencies, including with Executive Order 13286 (2003) the Department of Homeland Security. In fact, in a collaborative effort by the Homeland Security, Commerce, Interior, Transportation, Agriculture, and Defense Departments as well as NASA and the Environmental Protection Agency, the National Spatial Data Infrastructure put together a Geospatial One Stop Portal to provide access to geographic data (www.geodata.gov).

Finally, local data are important for economic and workforce development. For example, the National Neighborhood Indicators Partnership at the Urban Institute demonstrates how local data can be used to identify the dynamics of neighborhood change and to motivate local policy and program development. Moreover, the Urban Market Initiative at the Brookings Institution focuses on the uses of local information to facilitate the development of urban markets.

3. Data

3.1. Survey Data

The Decennial Census long-form survey has long been the primary source of detailed local data on place of residence, place-of-work, and journey to work. In concert with the introduction of the Census Transportation Planning Package (CTPP), which includes special tabulations aimed at summarizing key geographic and demographic characteristics of different areas for use in transportation planning, the Decennial began geocoding the place-of-work down to the block level in 1970. Other key information collected and tabulated included mode of travel to work, vehicles available, length of travel time to work, and the time of day when the person left for work. The CTPP has since been the main source of data for studying commuting patterns and their interaction with travel infrastructure demand and land use. Indeed, studies based on these rich, albeit often dated, datasets continue to be written in significant numbers (see, for example Shen 2000).

Major problems with using the Decennial Census to provide local data are its infrequent release dates (only once every ten years) and the point-in-time nature of its samples. To some extent these problems are being remedied with the rollout of the ACS, a national survey collecting information about the demographic, socio-economic, and housing characteristics of the population. The U.S. Census Bureau designed the ACS in part to produce current information about small geographic areas every year. The ACS takes the place of the long-form survey of the Decennial Census, and indeed the 60 questions that appear on the new survey closely resemble those that appeared on the Census 2000 long form. At present, about 250,000 addresses, or approximately one in 480 households, are in the ACS sample per month. Addresses in less-densely populated areas and small governmental units, such as American Indian reservations and rural towns and counties, are sampled at a disproportionately high rate. Additional information on ACS sampling procedures can be found on the Census Bureau's website (www.census.gov).

The ACS employs continuous measurement as a way of collecting data, conducting monthly household surveys and pooling the results to produce annual data. A major advantage of the continuous measurement approach is that the ACS can capture the seasonality of certain types of jobs and seasonal variation in demand for workers in certain regions; point-in-time surveys such as the Decennial, on the other hand, can result in some degree of misrepresentation in places of work, modes of travel, and population characteristics. Further, one can more accurately track changes in such variables over time using an annual survey based on continuous measurement; examining variation over ten-year intervals, as is required when using Decennial data, can obscure the timing of changes and delay the evaluation of policies, programs, or events.

While greater frequency, timeliness, and accuracy of place-of-work information are anticipated with the ACS and are welcomed by most geographic data users in the public and private sectors, numerous problems with these survey data persist. The sample is relatively small, encompassing about one in 40 households per year. Further, nonresponse to certain questions remains problematic. With respect to the place-of-work questions more specifically, nonresponse is likely nonrandom across individuals as well as industries and geographies. One might speculate that factors such as commuting modes, individuals' job stability and tenure, local population density, and other employer characteristics could affect respondents' ability to report accurate workplace addresses. Later in this article, we indeed find that nonresponse varies systematically over different observable characteristics of workers and their employers.

Moreover, in addition to small sample sizes and nonresponse, there is significant error in reporting. This type of error has been documented in the reporting of industry of work (Decressin et al. 2005) as well as in earnings (Roemer 2002). While misreporting of place-of-work is not well documented, our preliminary analysis suggests that errors in place-of-work reporting are as severe as, if not more severe, than errors in reporting other items.

Also, although place-of-residence geocoding has improved over time as a result of continuous corrections and updates to residential addresses, corresponding improvements in place-of-work geocoding have not kept pace. Indeed, the accuracy and completeness of place-of-work geocoding remains problematic not only because of definitional difficulties (such as the distinction between mailing and physical addresses of employers), but also because a place-of-work master list akin to that for place of residence does not exist.

The current approach to geocoding the place-of-work on the ACS at the U.S. Census Bureau involves three distinct processes: machine coding by the Geography Division, hand-coding by clerical staff, and hand-coding by specialists when the clerical staff refers an address to them. The Geography Division's system for geocoding involves first an automated address match to the Topologically Integrated Geographic Encoding and Referencing (TIGER) system database and an employer name match to an employer database. Such difficulties include, for example, the distinction between mailing and physical addresses of employers. A place-of-work response ungeocodable by the automated system is sent to the clerical operation, which uses the employer name and address reported in the survey, along with maps, the Internet, and other resources, for geocoding.

However, even after the Geography Division completes this process, some records still have missing values, and high geographic levels for these values such as county and place are allocated, or imputed. The allocation rate for a given item is calculated as the total number of responses allocated for that item divided by the total number of responses required, multiplied by 100. Item allocation rates for the 2000–2003 ACS are reported on the U.S. Census Bureau's web site (http://www.census.gov/acs/www/UseData/sse/index.htm). Item allocation rates for place-of-work and related statistics are in general slightly higher than those for responses to most questions regarding demographic characteristics on the ACS. For example, in 2003, 3.8% of state-of-work responses were allocated. For the county of work and Census Place-of-work, the percentages were even higher, 4.6% and 5.2%, respectively (for reporting of census data, Census Places include Census Designated Places, consolidated cities, and incorporated Census Places; there are some 7,000 Census Places in the U.S.). Those rates compare, for example, to allocation rates of 0.5% for telephone ownership, 1.0% for age, and 1.5% for marital status.

In our examination of the quality of the responses to the place-of-work questions using internal, unedited ACS 2003 microdata (making no distinction between response by mail-back, telephone, or personal interview; or between self-response and response by proxy), we found that respondents usually provide at least some information on their employers' locations. However, this information is often incomplete. For example, respondents sometimes provide no street information (ten percent of respondents); no city information (three percent of respondents); and, most frequently, no ZIP code (22% of respondents).

There is also evidence that respondents often misreport some geographic information about their employers. To the extent that workers in some industries and in some locations are not formally employed but rather work in the underground economy, place-of-work addresses are likely to be missing or misreported in a nonrandom fashion. Indeed, we find that workers in families at or below the poverty line as well as individuals employed in low-paying services industries and workers in very rural regions prove more problematic in terms of assigning place-of-work information. Such systematic misreporting can distort underlying distributions in the data, complicating imputation and potentially producing estimates that are unreliable at highly disaggregated geographic levels.

3.2. Administrative Data

Integration of survey and administrative data for the purpose of improving place-of-work information requires a record that contains information on both the employer and the employee. The records at the U.S. Census Bureau that provide this are the individual wage records from the unemployment insurance (UI) system of each U.S. state. These data are combined with establishment-level data, which are called the Quarterly Census of Employment and Wages (QCEW) data.

Given the sensitive nature of the dataset, it is worth discussing the confidentiality protection in greater detail. All data that are brought into the LEHD system have been anonymized in the sense that standard identifiers and names are stripped off and replaced by a unique "Protected Identification Key," or PIK. Only U.S. Census Bureau employees or individuals who have Special Sworn Status authorized to work on an approved project are permitted to work with the data, and they have not only been subject to a background check by the Federal Bureau of Investigation but also are subject to a $250,000 fine and/or five years in jail if the identity of an individual or business is disclosed. All projects have to be reviewed by the U.S. Census Bureau and other data custodians, and any tables or regression results that are released are subject to full disclosure review.

The LEHD Program currently has over 40 state partners. The data cover most of the period starting in 1991 and extend to at least early 2005 (while the precise start of data coverage varies by state, all LEHD states currently have data for the period 1998–2004). UI coverage is broad, covering over 95% of total wage and salary civilian jobs. Longitudinal (on a quarterly basis) job information on earnings and employers for the universe of covered workers in a state is also available in the UI data.

While UI coverage is broad and relatively comparable across states, UI wage records do not capture the following types of workers: federal employees, military personnel, self-employed individuals and independent contractors, and state residents who are employed out-of-state. Stevens (2002) provides a more detailed description of the workers who are and are not represented in the UI wage records.

The QCEW data are a major source of firm-side information. These data provide establishment-level characteristics, such as detailed industry codes, addresses, and ownership codes. Moreover, with few exceptions, the universe of firms in the QCEW micro-data contains the universe of firms covered by UI wage records. Thus, establishment-level characteristics are available for nearly all workers in the UI wage data.

The LEHD Program uses Group 1 commercial geographic coding software and the Census Bureau's Master Address File to map addresses into Census blocks and latitude/longitude coordinates. With the block-codes, entities are assigned to higher-level political and economic geographies, such as county, metropolitan area, voting districts, and so forth. Coordinates are used to calculate distance between the place-of-work and place of residence, and this distance measure is used in the production of a variety of public use data products.

Notably, for all states participating in the LEHD Program except Minnesota, UI wage records do not include an identifier for the physical establishment for workers at multiple-unit businesses; the only business identifier available is that of the parent firm. LEHD Program staff developed a multiple imputation methodology to assign a place-of-work to these individuals based on the size and hiring patterns of establishments within multi-unit businesses, and on the relationship between the place of residence of each worker and the location of each business. This imputation affects some 30% of workers, depending on the state under consideration (Abowd et al. 2005).

In investigating quality issues associated with business addresses in the administrative data, staff of the LEHD Program worked with the Bureau of Transportation Statistics and the states of Florida and Illinois (Almousa et al. 2003). They found that about 60% of establishments, covering about 70% of employment, had accurate geocoding on the physical location of establishments at the block level. One problem was that there was clear under-reporting of establishments within multi-unit firms, most often school districts and some large retail companies. As reported by Lane et al. (2003), however, during a one-year trial period, staff in Florida and Illinois was able to reduce the missing information by as much as 50%. The conclusions of the team were that the main quality problems can potentially be resolved in a number of ways, one of which included developing an ongoing cooperative relationship with states, particularly the local transportation agencies, to improve the QCEW. If local transportation agencies were to be routinely provided with updated origin-destination matrices, they would have strong incentives to improve the quality of the input information – and be best placed to do so.

In the same joint study between the LEHD Program and the Bureau of Transportation Statistics, the work and residence locations derived from LEHD data were compared with those derived from the CTPP using 1999 data for two geographic areas: Lake County, Illinois, and Miami/Dade Counties, Florida. The study concludes that LEHD and CTPP data are very similar for both states in terms of the place of residence. The density estimates are not only very similar, but the standard deviation of the estimates is quite close. In a recent article, Mix (2005) provides a detailed analysis of the differences between CTPP and LEHD data.

It is important to note that, despite growing demand for information on small area characteristics, the QCEW data have not been widely used for the production of statistics below the county level. The LEHD experience with its state partners, and in particular with states' labor market information directors (the custodians of the data), suggests that if products are produced that will help their clients, who include workforce investment boards, economic developers, and local transportation agencies, there is substantial potential and interest for them to enhance file quality.

4. Combining Administrative and Survey Data

There are two possible ways to use the LEHD data to improve place-of-work geocoding. One is to create a master business address list, similar to the Master Address File, and to match it to the information the respondent in the ACS provides. The second is to use the integration records to identify the employer of the ACS person and then to use the employer's address directly. In what follows, we describe both approaches in more detail.

4.1. Survey Response Method

The Survey Response (SR) method simply processes the reported place-of-work address through LEHD's master address list, the Geocoded Address List (GAL). A GAL has been created for each state. The GAL includes both business and residential addresses, including addresses from the ACS itself, each state's QCEW files, the U.S. Census Bureau's Business Register, and the U.S. Census Bureau's Master Address File.

The addresses are block-coded using Master Address File geocodes whenever possible, otherwise by Group 1 geocoding software. The GAL produces block codes at four levels of certainty, which ranked from highest to lowest certainty are the rooftop level, the block-face level, the block-group level, and the tract level. As we explain in more detail shortly, rooftop geocoding indicates known coordinates of the address, while the others indicate the geography at which we know the address exists (even if we do not know its precise latitude and longitude). In principle, rooftop and block-face levels of certainty are equally acceptable for block coding, as differences between locations along a block face do not affect the block code.

4.2. Employer Address Method

The Employer Address (EA) method is more complex than the SR method. Put simply, the approach consists of the following steps:

1. Determine the quarter containing the reference week of the journey-to-work questions asked in the survey. Because the ACS interview can occur on any day of the year, “last week” could be any week. Usually “last week” is the same quarter as the interview date, but not always. For example, if the interview was January 2, 2004 (the first quarter of 2004), “last week” was part of the fourth quarter of 2003.
2. Select the Protected Identity Key (PIK) of the respondent, which is a unique substitute for the Social Security Number.
3. Select the UI wage record to identify the firm at which earnings were highest (the dominant employer) during the reference quarter identified in Step 1, or the most recent quarter no earlier than one year before the reference quarter. If the dominant employer is a multi-unit, the EA method retrieves the first imputed establishment within the firm. Note that the ACS identifies only one employer per sampled person. This step makes no attempt to distinguish this employer from others existing in the UI data beyond the “dominant” employer distinction made here.
4. Select the geocode of the establishment from the GAL.

As with the SR method, the EA method generates geocodes at four levels of certainty, the highest of which is the rooftop level and the lowest of which is the tract level. We now turn to the results of applying these methods to enhancing the place-of-work information in the ACS.

5. Evaluation

The two approaches were applied for 23 states to three periods of ACS data: 2001, 2002, and the first month of 2004. The 23 states were CA, CO, FL, IA, ID, IL, IN, KS, KY, MD, ME, MT, NC, NJ, NM, OK, OR, PA, TX, VA, WA, WI, and WV. At the time of this analysis, only these states’ data had been fully processed by the LEHD Program, though many more states’ data were in process at the time and have since been incorporated. Moreover, at the time of this analysis, PIKs were not available for the 2003 ACS sample and were only available for the first month of the 2004 ACS sample.

5.1. Basic Results from the Two Methods

The results of the Survey Response (SR) method are in Table 1, which simply quantifies the accuracy of geocoding the place-of-work address reported in the ACS, by year, for all workers. Notionally, “rooftop” denotes confidence in the precise location of the address, “block face” means confidence in the correct side of the street between two corners, “block group” is a high level of certainty in the block group but not the block, and “tract” means the Census tract is certain but not the block within the tract. While the level of confidence varies across these categories, they all produce the full Census geocode: state, county, tract, and block. “Fail” means the geocoding system did not produce a block code, and of course blank addresses are impossible to geocode.

Table 1. Geocode Accuracy Using the Survey Response (SR) Method

		Percent of workers			Number of workers	
SR Quality	2001	2002	2004	2001	2002	2004
Rooftop	41.7	44.2	42.6	127,423	118,090	10,150
Block face	0.7	0.8	0.8	2,245	2,221	198
Block group	3.1	3.4	3.4	9,399	8,942	819
Tract	31.8	31.7	32.3	97,284	84,667	7,706
Fail	16.5	13.2	12.9	50,509	35,266	3,069
Blank address	6.2	6.6	7.9	19,022	17,691	1,892
Total	100.0	100.0	100.0	305,882	266,877	23,834

ACS 2001, 2002, and 2004 (one month), all workers.

The results are quite similar each year. The SR method is successful in geocoding between 41.7% and 44.2% of workers at the highest level of confidence (rooftop), about 4% at the block-face or block-group level of confidence, and between 31.7% and 32.3% at the tract level of confidence. The SR method fails to produce anything for between 12.9% and 16.5% of workers. The pattern seems to be that addresses are easy, difficult, or impossible to geocode, with little in between; as a result, the efficacy of the SR method may hinge on the acceptability of addresses geocoded only at the tract level of confidence.

The results of the Employer Address (EA) method, by year and accuracy category, are reported in Table 2. The EA method produces 26 possible outcomes. So many outcomes are possible because of the multi-dimensional source data; namely, because the QCEW may provide a physical address or mailing address, because the employing firm may be single-unit or multi-unit, and because there are four levels of certainty about the geocodes plus three methods of imputation.

Table 2. Geocode Accuracy Using the Employer Address (EA) Method

	Percent of workers			Number of workers		
EA Quality	2001	2002	2004	2001	2002	2004
Single-unit firm (43.5%)						
With physical address						
Rooftop	31.3	31.9	29.0	95,656	85,132	6,907
Block face	0.8	0.8	0.7	2,353	2,142	174
Block group	2.3	2.4	2.3	7,131	6,356	554
Tract	2.2	2.2	2.0	6,870	5,746	468
With mailing address						
Rooftop	2.1	2.4	2.4	6,555	6,378	581
Block face	0.2	0.2	0.3	480	498	68
Block group	0.6	0.6	0.6	1,759	1,620	138
Tract	0.1	0.1	0.1	364	304	22
County	2.6	2.9	3.0	8,093	7,623	719
Imputed county by industry	0.7	0.6	0.4	2,003	1,568	93
Imputed county without industry	0.2	0.2	0.2	646	469	52
Multi-unit firm (27.4%)						
With physical address						
Rooftop	18.4	19.3	18.0	56,282	51,572	4,298
Block face	0.6	0.6	0.6	1,929	1,720	137
Block group	1.9	2.0	2.0	5,862	5,470	470
Tract	2.2	2.2	2.0	6,852	5,793	468
With mailing address						
Rooftop	0.3	0.4	0.4	1,021	1,041	92

Block face	0.1	0.1	0.1	345	359	28
Block group	0.2	0.2	0.2	484	490	56
Tract	0.1	0.0	0.0	156	127	8
County	2.6	3.2	3.8	7,804	8,500	916
Imputed county by industry	0.2	0.2	0.1	482	428	36
Imputed county without industry	0.0	0.0	0.0	17	20	0
Employer missing (29.1%)						
Establishment missing	2.7	1.6	0.4	8,125	4,395	90
No in-range earnings	8.9	8.3	13.2	27,178	22,193	3,140
Firm missing	9.0	8.3	8.7	27,549	22,120	2,064
PIK missing	9.8	9.3	9.5	29,886	24,813	2,255
Total (100.0%)	100.0	100.0	100.0	305,882	266,877	23,834

ACS 2001, 2002, and 2004 (one month), all workers.

Not surprisingly, employer addresses from the QCEW seem generally easier to geocode than the place-of-work addresses reported in the ACS. The largest percentage of workers is at the rooftop level of confidence. Apart from workers for whom we find no employer, most of the EA results are single-unit physical address geocoded at the rooftop or multi-unit physical address geocoded at the rooftop.

A quality issue immediately arises from Table 2 in that, even after conditioning on individuals in the ACS who report working in one of the 23 LEHD states, 29.1% of ACS workers lack an employer in LEHD's database. There are several possible reasons for this.

5.1.1. No Establishment

For between 0.4% and 2.7% of workers, the EA method finds an employing firm, but the worker cannot be associated with an establishment within the firm.

5.1.2. No In-range Earnings

Timing issues result in a large number of workers lacking any record of earnings in the LEHD database in the 12 months preceding the ACS interview. This problem was particularly acute in 2004, when 13.2% of workers had no in-range earnings. At the time this table was generated in the fall of 2004, some states had not yet delivered 2003 data to LEHD, making it impossible to find the earnings and employer data within a year of the ACS interview.

5.1.3. No Firm

A sizeable portion of the sample, 8.3% to 9.0%, has no employer at all in the LEHD database. The failure to find a worker's employer seems to stem largely from the coverage of the QCEW and UI system. First, self-employed workers are categorically excluded from the UI system. Second, the UI system does not cover some government (especially federal) and agricultural workers. Third, people working without pay or in the informal economy will not appear in the UI system. Finally, there are a few non-workers in this group who are misidentified as workers in the ACS interview. The first two problems will be resolved as registers containing information about self-employed and federal workers are added to the LEHD database.

5.1.4. No PIK

Between 9.3% and 9.8% of workers in each year in the ACS lack a PIK, which renders it impossible to establish a link with administrative data. Improved PIKs will be possible in the future as the U.S. Census Bureau integrates the results of recent research to improve the process that produces the PIK. This should lower the proportion missing the PIK by a few percentage points.

5.2. The Potential Gains from Using Administrative Data

The potential gains from supplementing the current ACS geocoding system with the integration approach are summarized in Table 3. This table compares the effectiveness of the SR and EA methods with the current ACS geocoding systems in producing the block code, using alternative definitions of an acceptable LEHD result. The strict Definition 1 indicates a lower bound of how effective the integration approach could be, and the loose Definition 2 indicates an upper bound.

Table 3. Comparison of Geocoding Effectiveness by ACS Mode and Alternative Definitions of Acceptability

		Percent distribution of workers by LEHD method						Number of workers
		SR or EA	SR	EA	Reject	Fail	Total	
Definition 1:								
SR with certainty in the block or EA single-unit employer with certainty in the block								
ACS Mode	Machine	80.4	71.7	8.7	16.9	2.7	100.0	324,698
	Clerical/Referral	41.7	13.5	28.2	46.3	12.0	100.0	119,277
	Fail	31.0	7.5	23.5	52.2	16.8	100.0	152,618
	Total	60.0	43.6	16.4	31.8	8.2	100.0	596,593
Definition 2:								
SR with certainty in at least the tract or EA single or multi-unit employer with certainty in at least the tract								
ACS Mode	Machine	96.9	93.1	3.8	0.4	2.7	100.0	324,698
	Clerical/Referral	85.3	62.5	22.8	2.7	12.0	100.0	119,277
	Fail	80.1	60.5	19.7	3.1	16.8	100.0	152,618
	Total	90.3	78.6	11.7	1.5	8.2	100.0	596,593

ACS 2001, 2002, and 2004 (one month), all workers. Workers geocoded by both the SR and EA methods are tabulated in the SR column.

The strict and loose definitions produce a wide range of potential effectiveness. Under Definition 1, where the address is geocoded at the highest level of certainty using either the SR or the EA single-unit firm method, the integration approach would successfully geocode 41.7% of workers who would normally require clerical coding. Of these workers, 13.5% would be geocoded by the SR method and 28.2% by the EA method. The integration approach would also block-code 31.0% of workers for whom the current ACS geocoding systems fail to produce a block code.

A much less strict option (Definition 2) accepts all the LEHD geocodes except the imputations to county centroids and indicates an upper bound of the integration approach's effectiveness. It may compromise quality, but this definition dramatically relieves the clerical workload, and fills in four-fifths of the data currently missing. Under Definition 2, the integration approach successfully geocodes 85.3% of workers currently block-coded by the clerical operation (62.5% by SR and 22.8% by EA), and fills in 80.1% of blanks.

5.3. Further Investigation of Quality Issues

A number of quality issues arise from the combination of different data sources. The SR approach supplements the ACS data with one additional source, and the EA approach supplements it with two. This section summarizes the results of an analysis of the issues surrounding the integration of different data sets.

5.3.1. Response Error

One possible concern with the SR approach is related to the possible misreporting of employer ZIP codes in the ACS. LEHD's geocoding system relies heavily on the ZIP code and typically results in a geocode confident only in the tract when the street portion of the address is deficient. This reliance on the ZIP code may cause "heaping" of workers into certain tracts, which we will attempt to detect and mitigate later in this article. One might suspect as well that a respondent uncertain of the employer's ZIP code might guess in the survey that it is the same as the home ZIP code, which would lead to a place-of-work geocode that wrongly places the employer in the same tract as the residence. However, closer examination of the data suggests that this is not likely to be a major issue; indeed, those for whom we geocoded place-of-work with confidence in the tract are no more likely (15.6%) to work in their home tract than workers for whom we geocoded place-of-work with confidence in the rooftop (17.6%).

5.3.2. Validity

The first major issue with the EA approach is that the ACS respondent may not work where the employer is located. In order to investigate this, we compare the rate of agreement between the EA geocode and the SR geocode, where the SR geocode is at the highest level of certainty. Essentially, in these cases, we know where the person works because the address provided in the ACS was easily and precisely geocodeable; it can therefore serve as a standard to which we can compare the EA geocode. The results of this analysis are presented in Table 4 (notably, for reasons discussed in Stevens (2002), accurately geocoding establishments affiliated with multi-unit firms as opposed to single-unit establishments can be more difficult, and therefore a disproportionate number of individuals employed at multi-unit firms are excluded from the figures).

An examination of the first row shows that 61.8% of workers at single-unit firms that provided a physical address that could be geocoded at the highest level of certainty have an EA-based block in perfect agreement with the SR-based block. This seems to indicate that most workers do, in fact, work at the addresses identified in the administrative files. The right-most column of the first row shows that almost 70% of the rooftop single-units are consistent at the tract level with the rooftop addresses reported in the ACS. Among the remaining 30% that are in a different tract by EA, about one-half are within five miles of the SR address. Even imputed establishments (those within multi-unit firms) are fairly consistent with SR; more than 40% geocoded at the rooftop agree with the SR tract. It is possible that establishments missing from multi-unit firms in the QCEW play a role in lowering the consistency between the EA and SR methods.

Table 4. Consistency of Results by Method

EA	Percent distribution of workers by lowest level of agreement between EA and Rooftop SR: EA missing	State	County	Tract	Block Group	Block	Total	Number of workers	At least the tract
Single-unit firm									
With physical address									
Rooftop	0.0	10.1	21.1	2.4	4.6	61.8	100.0	98,294	68.8
Block face	0.0	16.9	41.6	5.9	9.8	25.8	100.0	1,554	41.6
Block group	0.0	22.8	54.2	11.4	8.4	3.2	100.0	3,014	23.0
Tract	0.0	27.1	57.1	7.2	6.2	2.4	100.0	3,611	15.8
With mailing address									
Rooftop	0.0	10.7	19.7	1.9	4.6	63.1	100.0	6,938	69.6
Block face	0.0	10.4	25.2	4.3	16.1	44.0	100.0	461	64.4
Block group	0.0	17.1	56.5	14.7	9.5	2.3	100.0	1,020	26.5
Tract	0.0	21.7	57.9	8.1	5.4	6.8	100.0	221	20.4
County	100.0	0.0	0.0	0.0	0.0	0.0	100.0	7,673	0.0
Multi-unit firm									
With physical address									
Rooftop	0.0	33.6	25.9	1.9	4.3	34.3	100.0	51,541	40.5
Block face	0.0	42.7	32.2	4.1	6.8	14.3	100.0	1,254	25.1
Block group	0.0	42.0	43.8	5.8	6.2	2.3	100.0	3,156	14.3
Tract	0.0	42.7	47.1	5.1	4.0	1.1	100.0	3,859	10.2
With mailing address									
Rooftop	0.0	27.7	27.9	2.7	4.7	37.0	100.0	1,003	44.4
Block face	0.0	23.5	26.2	4.8	16.1	29.4	100.0	378	50.3
Block group	0.0	21.1	62.7	7.3	6.7	2.1	100.0	327	16.2
Tract	0.0	39.3	41.1	10.7	6.3	2.7	100.0	112	19.6
County	100.0	0.0	0.0	0.0	0.0	0.0	100.0	6,435	0.0
Firm or establishment missing	100.0	0.0	0.0	0.0	0.0	0.0	100.0	64,812	0.0
Total								255,663	

ACS 2001, 2002, and 2004 (one month), workers geocoded at the highest level of confidence by the SR method.

A potential related problem is due to differences in mailing and physical addresses. Indeed, while the aim of the ACS is to collect address information on the physical location where a worker went to work, the QCEW file often includes only an employer's mailing, business office, headquarters, or other address, which may not be where a worker actually works. However, this analysis reveals that single-unit firms that provided only a mailing address are just as consistent with the SR-based block as those that provided a physical address; the SR block and EA block agree perfectly for 63.1% of workers among mailing addresses and 61.8% among physical addresses. This could occur because ACS respondents feel obliged to provide an address and the only one available is a mailing address, or because employers report a physical address in the mailing address field of the QCEW, or for other reasons.

There is a high level of agreement between the geocodes derived from the two LEHD methods. Further investigation shows that in general, when the address reported in the ACS and the employer-based address are similarly geocodeable, the two methods identify the same physical location, suggesting that the address provided by the QCEW is, by and large, the same as the one reported in the ACS.

Despite these encouraging results, it is important to note that a substantial number of geocodes we assign to individuals' workplaces may nevertheless be invalid. We can to some extent assess validity by exploiting systematic variation in the amount of discrepancy by industry between the geocodes produced by the SR and EA methods. Our examination of the data suggests that discrepancies are more likely to occur among workers in certain industries; indeed, among the places of work we geocode at the highest level of certainty by the SR approach and the single-unit EA approach, certain industries disproportionately produce a discrepancy in the workers' block, tract, or county level geocode. For example, real estate is an industry in which it is easy to imagine someone working at a location different from the address appearing in an administrative file, and indeed we find (in unreported results) that individuals employed in real estate as well as other "fragmented" industries such as educational services have higher levels of disagreement in geocodes produced by different methodologies.

Another way to investigate whether an employer address is a good substitute for an address reported in the ACS interview is to compare the two commuting distances derived from them. Table 5 reports the distance from the centroid of each worker's residence block to the place-of-work address as derived from the SR and EA methods. Including only workers we successfully geocode by both methods in the table minimizes the differences in universe coverage.

Table 5. Commuting Distance by Place-of-Work Geocoding Method

	Percent distribution of workers by commuting distance (miles) 0 to 2.5	2.5 to 10	10 to 20	20 to 50	50 to 100	100 or more	Total	Number of workers	Percent distribution by quality
By SR Quality									
Rooftop or Block Face	29.2	42.4	19.3	8.1	0.6	0.4	100.0	194,069	56.5
Block group or Tract	33.1	40.9	17.1	7.4	0.9	0.6	100.0	149,575	43.5
All SR	30.9	41.8	18.4	7.8	0.7	0.4	100.0	343,644	100.0
By EA Quality									
Single-unit firm									
Rooftop or block face	25.1	40.7	19.1	9.6	2.2	3.3	100.0	170,240	49.5
Block group or tract	21.3	41.1	19.7	10.2	2.8	5.0	100.0	25,709	7.5
Multi-unit firm									
Rooftop or block face	14.3	32.2	18.9	15.7	9.4	9.5	100.0	96,285	28.0
Block group or tract	13.8	33.4	20.3	15.7	8.3	8.5	100.0	21,036	6.1
SU or MU, county	4.8	27.1	22.1	18.3	10.4	17.2	100.0	30,374	8.8
All EA	19.3	36.7	19.4	12.5	5.4	6.7	100.0	343,644	100.0

ACS 2001, 2002, and 2004 (one month), workers with SR and EA geocodes, excluding 389 workers missing place-of-residence geocodes.

The EA method produces longer estimated commutes, especially among multi-unit firms. According to the EA method, 12.5% of workers travel 20 to 50 miles to work, while according to the SR method, only 7.8% of the same workers travel this far. This phenomenon raises concern about the EA method producing upward bias in the distance traveled. One possible cause is that workers are heaped into a geographic area when a multi-unit firm appears as a single-unit establishment or fails to report all its establishments in the QCEW. In these instances, the workers' place-of-work is imputed to be one of the reported establishments, one that might be located substantially farther away from individuals' residences than is their true workplace. Certainly, extremely long commute distances produced by the EA method are too numerous to ignore. The EA method produces 15 times more commutes over 100 miles than the SR method.

A final method of evaluating the validity of an employer address is simply to inspect it clerically side-by-side with the address reported by the respondent in the ACS. Our limited clerical review found frequent discrepancies; however, it carried an extraordinarily high time cost, and it was unclear in the majority of cases which process produced a more accurate geocode.

5.4. Addressing the Quality Issues

Discrepancies in geocodes motivate developing a set of filtering rules for place-of-work information produced by the SR or the EA method. For each worker's place-of-work, we select the best available geocode, applying filters to minimize the data quality problems inherent in the results of the SR or EA method. These rules are quite aggressive in limiting which geocodes are acceptable. We consider the quality of each geocode obtained from the SR or EA method, accepting only those that satisfy the rules set out below. If a geocode described in Rule 1 does not exist, we look for the next most preferable, a geocode meeting the conditions listed in Rule 2; if this does not exist, then we proceed to the next most preferable, Rule 3, and so forth, for each worker.

1. Block-confident SR.

2. Block-confident single-unit EA, where
    - The number of EA workers is less than 133% of SR workers in this tract, among workers geocoded by both methods (Heaping Filter).
    - The industry is not one that disproportionately produces a discrepancy in block, tract, or county coding among workers with block-confident SR and block-confident single-unit EA geocodes (Industry Filter). These industries consist of Educational Services; Administrative, Support, and Waste Management Services; Public Administration and Active Duty Military; and Real Estate and Rental and Leasing.
    - The place-of-work is less than 100 miles from the residence (Distance Filter).
3. Tract-confident SR, less than 100 miles from the residence (Distance Filter).
4. Other EA passing through the Heaping, Industry, and Distance Filters, and excluding county centroids.
5. Any SR if the SR block and EA block agree, or are within two miles of each other.

Thus, for each geocode obtained from the SR or the EA method, the first filter identifies whether that geocode was produced by the SR method with block confidence. If it was, it is retained. If not, the second filter is whether the geocode was produced by the EA single-unit method with block confidence, where the heaping, industry, and distance filters are also satisfied. Again, the address is kept if this condition is satisfied. Similar steps apply when it comes to conditions three through five. Note that the geocode in condition five is accepted because two independent methods have produced exactly or nearly the same result at the block level. In this case, it seems sensible to accept the geocode as true regardless of the filters or the precision of geocoding.

Table 6 demonstrates the impact of applying these rules, quantifying the percent of workers for whom the integration approach produced a place-of-work block code, by the geocoding system currently used to produce the block code (machine or clerical/referral). The overall efficacy of the integration approach with the filtering rules described above, 84.0%, is closer to that which we observed under the loose definition of acceptable quality (90.3%) than under the strict definition (60.0%) in Table 3. Moreover, using both methods and the filtering system successfully geocodes the place-of-work for 72.8% of ACS workers that normally would have been clerically geocoded, and 78.9% of the places of work not block-codeable by any of the current ACS methods. The integration approach even enabled block-coding 23.3% of workers for whom no place-of-work information at all was reported in the survey.

Table 6. Efficacy by Current ACS Mode

	Percent of ACS Workers				
	ACS All	ACS Machine	ACS Clerical/ Referral	Non-blank Address, ACS Fail	Blank ACS Address
Geocoded	**84.0**	**94.8**	**72.8**	**78.9**	**23.3**
SR Total	**74.0**	**90.1**	**55.6**	**64.9**	**0.0**
Block-confident SR	43.6	71.7	13.5	9.0	0.0
Tract-confident SR, filtered	30.4	18.4	42.1	55.9	0.0
EA Total	**10.0**	**4.7**	**17.2**	**14.0**	**23.3**
Block-confident SU EA, filtered	7.4	3.8	12.3	10.8	13.2
Tract-confident SU EA, filtered	0.5	0.1	0.8	0.9	2.7
Block-confident MU EA, filtered	1.6	0.6	3.3	1.8	5.8
Tract-confident MU EA, filtered	0.4	0.1	0.8	0.6	1.6
Restored	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
Suppressed	**6.3**	**2.1**	**12.5**	**7.8**	**23.7**

EA Heaping filter	3.5	1.2	7.0	4.0	12.8
EA Industry filter	1.8	0.6	3.7	2.2	6.9
Distance filter	1.0	0.3	1.8	1.5	4.0
Fail	**9.7**	**3.1**	**14.7**	**13.3**	**53.0**
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Total Number	**596,593**	**324,698**	**119,277**	**127,232**	**25,386**

ACS 2001, 2002, and 2004 (one month), all workers. "Blank ACS Address" means blank street address and blank ZIP code. Filters are applied to a worker in sequence; only the last filter that operated is tabulated.

5.5. Analysis of Impact

Integrating administrative data with survey data clearly has its advantages with respect to improving the comprehensiveness and accuracy of place-of-work information. However, especially when it comes to public use of the data, representativeness is of the utmost importance, and on this dimension the integration approach yields tangible benefits. In an effort to assess the bottom line with respect to the net impact of combining existing methods with LEHD's approach to integrating administrative data, we conducted several exercises comparing the edited pooled 2001 and 2002 microdata produced by the current system (i.e., data that incorporate machine, clerical, and referral geocodes) with the enhanced pooled 2001 and 2002 microdata supplemented with administrative data (i.e., data that incorporate the clerical and referral geocodes and, when those are missing or incomplete, geocodes supplied by the SR and EA system using the filtering system described above). In order to ensure that the two datasets are comparable, we first restrict attention to only those workers in the ACS who are working, who have a valid place-of-work state, and who work in one of the 23 states for which we supplement the data with administrative information.

The enhanced data show greater geographic dispersion in economic activity than the unenhanced files. Within the 23 states under consideration, 17% more blocks and 6% more tracts are identified as ones on which people work when administrative information on employer addresses is used to supplement the survey data. The increase in the number of blocks identified as the work site for at least one worker is greater than 25% in some of the smaller, more rural states. These patterns hint at underlying differences in the degree of incompleteness in survey-based place-of-work information for workers in urban as opposed to rural areas. As Table 7 reveals, the enhanced data enable us to identify the precise location of more rural economic activity, a significant share of which cannot be accurately geocoded in the current system. Whereas roughly 80% of workers who live in metropolitan areas have block-level place-of-work geocodes in the edited (but unenhanced) data, only about 60% of workers in non-metropolitan areas do. Augmenting the data with administrative information nearly equalizes the proportion missing across these categories, increasing the proportion of workers with block-level geocodes in each to nearly 95%.

Finally, there is a substantially greater gain in identifying the place-of-work for workers who live in households in poverty than other workers. The U.S. Census Bureau uses a set of money income thresholds that vary by family size and composition to determine poverty status. Supplementing the current system of geocoding place-of-work with the SR and EA methods, the proportion of workers in poor households with block-level place-of-work geocodes increases by almost 25 percentage points, compared with an increase of under 20 percentage points for workers in households above the poverty line.

Table 7. Comparison of Geocoding Rates

	Percentage with Block-Level Geocodes (Pooled 2001-2002 ACS Data, Using Person Weights)	
	Edited ACS	Edited & Enhanced ACS
All Workers	76.1	95.1
Place of Residence		
Central City in MSA	80.4	94.5
Non-Central City in MSA	79.9	95.4
Non-MSA	59.3	95.1
Poverty		
Not in Poverty	76.7	95.3
In Poverty	66.7	91.6

ACS 2001 and 2002, all workers.

While differences at higher levels of geography than the block level are not as pronounced, the added coverage and representativeness the LEHD approach affords at the subcounty level are substantial. Given the strong and growing demand for local data in both the public and private sectors, the availability of a comprehensive and representative sample that accurately places workers and the establishments at which they are employed geographically is becoming all the more important.

6. Conclusions and Next Steps

We began this analysis by noting the substantial demand for more detailed local information and the gains that could be derived from combining administrative with survey data. Our findings using an administrative dataset based on U.S. state UI records to supplement ACS data suggest that, indeed, there are likely important gains that NSOs may reap by following such an approach. Using the strictest possible definition of an "acceptable" geographic code, this approach would successfully geocode 41.7% of workers who would normally require clerical coding, and geocode 31.0% of workers whom the current ACS geocoding systems fail to geocode. Under careful but less stringent definitions, the approach would successfully geocode 78.9% of workers currently block-coded by the clerical operation, and would also fill in 72.8% of blanks. In other words, using administrative data to correct for missing or incomplete geographic information in survey data could generate nontrivial cost savings for NSOs.

Furthermore, using the current geocoding system, involving machine, clerical, and referral work, and merely supplementing it with processes incorporating administrative data would not only substantially increase coverage, but would also improve it for particular groups such as rural and disadvantaged populations. Better information on the place-of-work for workers in these populations is of particular use to policymakers in addressing spatial mismatch issues and implementing targeted welfare programs. The improvements in coverage and representativeness similarly increase the reliability of local area statistics on land-use, commuting patterns, and clusters of economic activity, all of which are important in planning for economic development, transportation infrastructure, and emergency preparedness.

Although there are clear quality issues, our analysis suggests that a combination of technical filters and feedback from users is likely to mitigate the problems in the long run substantially. In addition, the expansion of the administrative records to include the self-employed and federal workers will improve the universe coverage.

Finally, while this study has focused on the way in which administrative records can enhance survey responses, the links that the LEHD Program has created make it possible for survey data to enhance administrative records. LEHD's administrative data have already been used to create origin-destination matrices at the block-group level, so that users can map the flows of workers to work from a given area (commute sheds) and the flows of workers to work to a given area (labor sheds). The ACS, which provides household information together with commuting details such as time of day and mode of transport, could be combined with administrative records to model these flows in an even richer fashion.

7. References

Abowd J., Stephens, B., and Vilhuber, L. (2005). The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. U.S. Bureau of the Census, LEHD Working Paper.

Almousa, W., Foster, G., Lane, J., Mix, W., O'Connell, M., Putnam, G., and Roemer, M. (2003). An Evaluation of the Use of LEHD Data for Transportation Planning. U.S. Bureau of Transportation Statistics, LEHD Staff Report.

Andersson, F., Lane, J., and McEntarfer, E. (2004). Successful Transitions out of Low-Wage Work for Temporary Assistance for Needy Families (TANF) Recipients: The Role of Employers, Coworkers, and Location. The Urban Institute, Prepared for the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Contract No. HHS-100-01-0010, UI Project No. 07245-003-00.

Andersson, F., Holzer, H., and Lane, J. (2005). Moving Up or Moving On: Who Advances in the Low-Wage Labor Market? New York: Russell Sage Foundation Press.

Decressin, A., Lane, J., McCue, K., and Stinson, M. (2005). Employer-Provided Benefit Plans, Workforce Composition and Firm Outcomes. U.S. Bureau of the Census, LEHD Technical Paper TP-2005-1.

Executive Order 12906. (April 13, 1994). Federal Register, Volume 59, Number 71, 17,671-17,674.

Executive Order 13286. (March 5, 2003). Federal Register, Volume 68, Number 43, 10,619-10,633.

Horner, M.W. (2004). Spatial Dimensions of Urban Commuting: A Review of Major Issues and Their Implications for Future Geographic Research. The Professional Geographer, 56, 160-173.

Lane, J., Roemer, M., Mix, W., Putnam, G., Almousa, W., O'Connell, M., and Foster, G. (2003). An Evaluation of the Use of LEHD Data for Transportation Planning. Final Report to the Bureau of Transportation Statistics.

Lomax, T. and Schrank, D. (2005). The 2005 Urban Mobility Report. Texas Transportation Institute, Texas A&M University.

Mix, W.A. (2005). Evaluating the Local Employment Dynamic Program as an Alternate Source of Place-of-work Data for Use by Transportation Planners. Working Paper, Buffalo State College, Department of Geography and Urban Planning.

Pugh, M. (1998). Barriers to Work: The Spatial Divide between Jobs and Welfare Recipients in Metropolitan Areas. Brookings Institution Center on Urban and Metropolitan Policy. Washington, Brookings Institution.

Richardson, C. (June 28, 2005). Introducing Google Earth. Webpro News. http://www.webpronews.com/insiderreports/searchinsider/wpn-49-20050628IntroducingGoogleEarth.html

Roemer, M. (2002). Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation. U.S. Bureau of the Census, LEHD Technical Paper TP-2002-22.

Shen, Q. (2000). Spatial and Social Dimensions of Commuting. Journal of the American Planning Association, 66, 68-82.

Stevens, D. (2002). Employment that is not Covered by State Unemployment Insurance Laws. U.S. Bureau of the Census, LEHD Technical Paper No. TP-2002-16.

Terdiman, D. (2005). Google Gets Earthy. Wired News. June 30. http://www.wired.com/news/technology/0,1282,68042,00.html