

An Evaluation of the Use of LEHD Data for Transportation Planning

Final Report
September 16, 2003

Julia Lane and Marc Roemer
LEHD Program
US Census Bureau

Wende Mix
Westat

George Putnam and Waleed Almousa
Illinois Department of Employment Security

Mike O'Connell and George Foster
Florida Agency for Workforce Innovation

Contents

Executive Summary.....2
Background.....3
Data Description.....4
Confidentiality Protection.....5
Improvements to Establishment Address.....6
Comparison of LEHD data to CTPP.....17
Lessons Learned.....20
Next steps.....21

Appendices

1. Palm Beach Report
2. Florida Summary Report
3. Report on Florida Priority Files
4. Geocoding Project Documentation (Illinois)
5. Place of work imputation for multi unit businesses
6. Noise addition approach to disclosure proofing

Executive Summary

The LEHD program has created a new database infrastructure that could, in principle, be useful for transportation planning. This report documents the results of a one year analysis of the LEHD data in Florida and Illinois for this purpose. Its main findings are

- The LEHD data can be used to provide transportation planners with information on commuting patterns on a regular basis at the block level, together with information on the characteristics of both those blocks.
- The place of work information has quality problems in two main areas: missing information on the physical location of establishments, and under-reporting of multiple units. The LEHD state partners were able to reduce the missing information by as much as 50% - an ongoing program would continue to close the gap.
- The main quality problems can potentially be resolved in a number of ways
 - Developing an ongoing cooperative relationship with states – particularly the local transportation agencies to improve the ES202 (establishment file). If local transportation agencies are routinely provided with updated Origin-Destination matrices, they will have strong incentives to improve the quality of the input information – and be best placed to do so.
 - Matching the ES202 data to the Census Business Register and the Census of Governments, to improve the quality of the physical addresses and the multi-unit breakout information.

Another approach that might be useful for transportation planning is to leverage both the LEHD data and ACS data by integrating the two data sources to model commuting modes and the time of day of the commute.

Background.

The Bureau of Transportation Statistics (BTS) established a partnership with the Longitudinal Employer-Household Dynamic (LEHD) Team and two of its state partners – Florida and Illinois - in order to examine the potential to develop transportation specific data from the LEHD data holdings.

The LEHD program began in FY1999 and received its first data files from state partners in the winter of 2000. As of September 2003, the following twenty-seven states are the official LEHD state partners supporting the project: CA, FL, IL, MD, MN, NC, NJ, OR, PA, TX CO, ID, KS, MT, MO, NM, WV, VA, DE, GA, IA, KY, MI, ND, OK, WA, and WI. The first 18 have had data processed; the LEHD program is awaiting funding before processing the remaining nine.

There were three key deliverables:

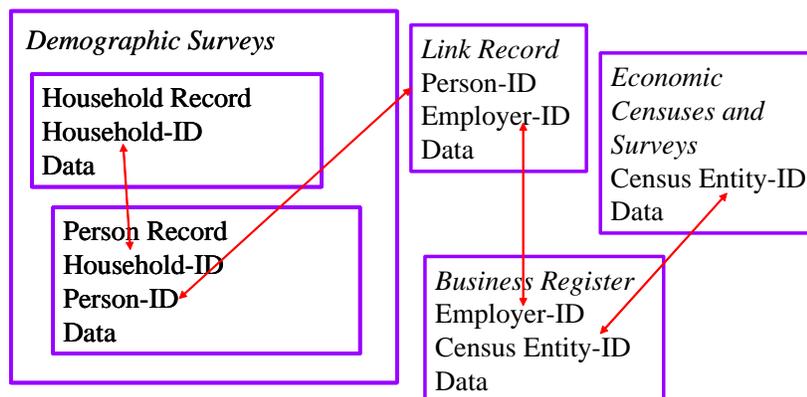
1. Origin/Destination (O-D) employee numbers from household to place of employment.
2. Information on the characteristics of workers by block residence
 - number of workers living on each block
 - proportion of workers earning low, medium, or high annual wages
 - mean annual wages
3. Information on the characteristics of businesses by block
 - mean quarterly pay per worker
 - industries operating on each block (SIC, Standard Industrial Classification division)

In addition, the LEHD program agreed to work with FL and IL to improve the quality of the place of work coding on the ES202 data and report on the value added of different approaches.

Data Description

The LEHD database enables us to match workers with past and present employers, together with employer and worker characteristics (Abowd, Lane and Prevost, 2000). This database consists of quarterly records of the employment and earnings of almost all individuals from the unemployment insurance systems of a number of US states in the 1990s¹. The basic concept underlying LEHD is illustrated in Figure 1. The link record is the Unemployment Insurance wage record data for each state which consists of almost the full universe of employers and workers every quarter². The Unemployment Insurance records have also been matched to internal administrative records containing information on date of birth, place of birth, race, and sex for all workers, thus providing limited demographic information.

The Longitudinal Employer - Household Dynamics Program



The value of these data to the transportation community lies in the fact that the data can be matched to place of residence and place of work information – in principle for all workers in the dataset. The place of residence of each worker (from 1999 on) is derived from an extract from the Census Bureau’s Statistical Administrative Records System (StARS), which incorporates data from the Internal Revenue Service (IRS) 1040 and 1099, Housing and Urban Development (HUD), Medicare, Indian Health Service, Selective Service System to get a “best” residential address. The match rate to the LEHD data exceeds 90%. The quality of the place of work information is less impressive – and is the subject of a subsequent section.

The differences between the job-based UI wage record data and the worker based data with which many researchers are familiar means that some key definitional decisions need to be made with regards to earnings and employment. In particular, because all jobs held by all workers are

¹ Because of the sensitivity of these data it is worth noting that the data are anonymized before they are used in any Census Bureau projects. Any research that is engaged in must be for statistical purposes only, and under Title 13 of the U.S. code, any breach of confidentiality can result in prosecution in which violators are subject to a \$250,000 fine and/or 5 years in jail.

² Stevens (2002) describes coverage issues related to the LEHD database.

in the data-set, it is possible to analyse two different facets of the labor market – both jobs and employment. The two obviously differ to the extent that there is multiple job holding, and to the degree in which there is churning of workers through different sets of jobs. In this pilot analysis we chose to focus on one employer-employee match – the employer is taken to be that employer from which an employee derived his/her maximum earnings in the second quarter of the reference year. However, the earnings definition that is chosen is the total earnings from all sources during the year.

A second important issue is that Unemployment Insurance wage records – except for Minnesota - do not include the physical establishment for workers who work for multi-unit businesses – only the identifier for the multi-unit. 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 the multi-unit business, and the relationship between the place of residence of each worker and the location of each business. This procedure is discussed in appendix 5.

Confidentiality Protection

The Bureau of Transportation Statistics specifically requested the release of data at the block level. The Census Disclosure Review Board approved the release, conditional on the following restrictions.

Block-level origin-destination data:

- 1) The origin block (or groups of blocks) must have at least five persons residing at that block.
- 2) There must be at least 3 distinct destinations for each origin.

The effect of these restrictions was to reduce the number of blocks that could be released for Florida from 229,752 to 166,240 – about 72% of all blocks and 96.8% of employment.

Information on the characteristics of workers by place of residence

1) Number of workers with low earnings (under \$12,000 annual earnings); high earnings (over \$35,000 annual earnings); and medium earnings (between \$12,000 and \$35,000). These statistics were generated from a kernel density estimate (using Silverman's rule-of-thumb and 10,000 bins) of the distribution of earnings for each block for which there are more than five workers. The input values are based solely on the actual maximum and minimum observed value of earnings for the workers reside on the block. The resulting counts are thus derived from a completely synthetic earnings distribution.

- 2) Average earnings of workers was also generated from a kernel density estimate

Information on the characteristics of businesses by block

- 1). The list of the industry divisions of businesses on that block (1987 SIC Divisions A - J.) is not disclosure protected.
- 2). The average payroll/employee of businesses on that block (i.e. sum of total payroll for all businesses on block, divided by total employment on the block). The numerator was protected by the noise addition methodology described in Appendix 6.

Improvements to Establishment Address

Approximately 60% of the ES202 data (representing over 60% of employment in the two states, have valid (priority 0) physical addresses.³ Our goal was to evaluate the level of effort required to provide accurate information on where people work within the framework of the ES202 data program. Ultimately, we wanted to determine the viability of using ES202 information in building home to work trip tables for transportation planning analysis.

We focused on how many invalid addresses could be located and at what cost. The effort was led by Westat, in coordination with Florida and Illinois Employment Security Staff. The Employment Security Staff in each state also received assistance from staff at the Metropolitan Planning Offices (MPOs) local transportation agencies as well as the State Departments of Transportation. The flow of work is described in Figure 2

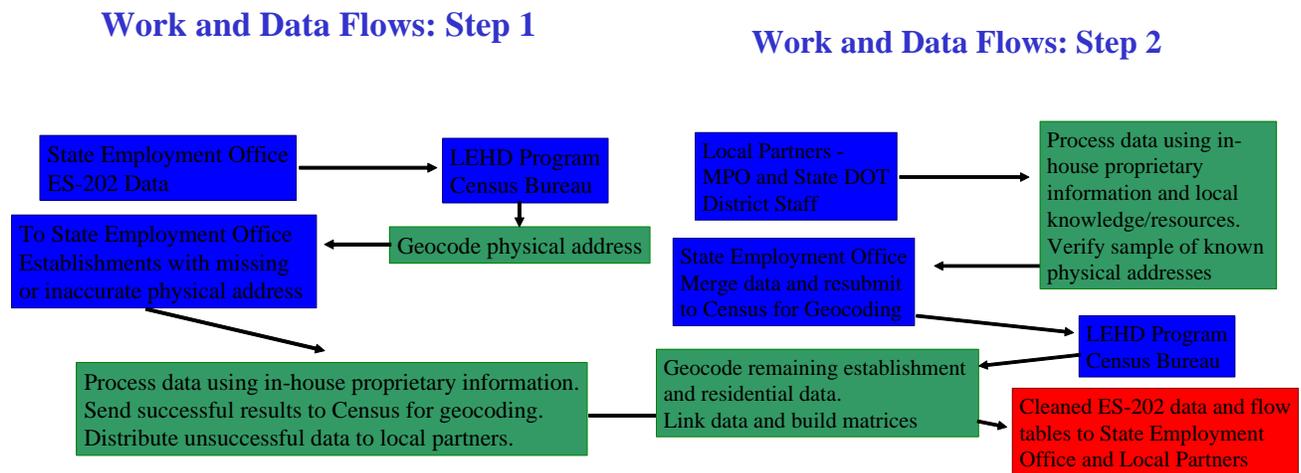


Table 1 shows the number of establishments and employment in the ES202 files by the priority. The efforts by the states focused on the Priority 2 records, those establishments without a physical address that would geocode to an acceptable level.

Priority	Florida Number of Establishments	Illinois Number of Establishments
0	250,035	204,975
1	19,795	-
2	114,658	142,363
TOTAL	384,488	347,338
Priority	Florida Total Employment	Illinois Total Employment
0	4,075,597	4,341,350
1	502,427	-
2	2,002,997	1,773,886
TOTAL	6,581,021	6,115,236

Table 1: Summary of 2001 ES202 data.

³ An address in Illinois was invalid if the physical address did not geocode to the block face, or if the physical address was unavailable, or if there was a Code1 failure, or if the address was out-of-state. An address in Florida was invalid if the physical address was unavailable, or if there was a Code1 failure, or if the address was out-of-state.

Table 2 shows the proportion of establishments with valid addresses. Also provided is the proportion of Employees associated with these establishments. In Illinois, over 70% of the employees are associated with establishments with valid addresses.

Priority 0	Florida	Illinois
Establishments	65%	59%
Employment	62%	71%

Table 2: Establishments with valid physical addresses.

Data Cleaning

Florida establishments were separated into six geographic areas to facilitate the needs of the different groups who would be working on the files. Broward, Miami-Dade, and Palm Beach counties were separated from the other counties that made up District 4 of the Florida Transportation Department. The MPO and District 4 staff worked on these files. The Employment Security Staff focused on the remaining establishments. Table 3 shows the number of establishments and the employment associated with these establishments by area.

AREA	Look-up Establishments	Total Employment
Broward	12520	186104
Miami-Dade	18735	320417
Palm Beach	9134	128434
District 4	2519	34680
Rest of State	71750	1333362

Table 3: Florida Summary by Area.

The Illinois ES202 data separated the state into 11 regions, namely Cook, DeKalb, DuPage, Grundy, Kane, Kankakee, Kendall, Lake, McHenry and Will Counties plus “all other counties”. Table 4 provides a summary of the total establishments and the total employment within each of these areas. Cook County and the Other county area account for more than two thirds of the records and employment.

The Employment Security Staff in Illinois initially focused on Cook Co. They split this data into three sets. Employment Security Staff worked on multi-units and large employers. Chicago Area Transportation Study (CATS) staff worked on establishments with 5 to 100 employees and Westat worked on 20,000 very small establishments. Later, the Employment Security Staff recruited the state DOT and several MPOs to work on establishments in their jurisdictions.

AREA	Priority 2 Records	Total Employment
Cook	43894	631083
DeKalb	794	12430
DuPage	12389	136997
Grundy	431	4289
Kane	3968	42697
Kankakee	846	11452
Kendall	526	7091
Lake	7176	91722
McHenry	2693	22184
OtherCty	65780	774037
Will	3866	39904

Table 3: Illinois Summary by Area.

Illinois analysis differed from Florida somewhat in that some attention was paid to SIC codes to aid the investigation. SIC codes identify the type of establishment and can be used to sort and target particular establishments. For instance, restaurant guides may be used to locate eating establishments (SIC 5812). Also, establishments that are less likely to have home to work commuters, for instance SIC codes associated with contract work - drywall, carpet cleaning, floor installation, special trade contractors, etc. or private households, can be lowered in priority for look-up. Data were sent to Westat as the files were processed. Documentation from some of the participants on their experience is included as appendices 1-4.

Characteristics of look-up data:

The vast majority of the establishment records in the files sent to Florida and Illinois represent small employers, under 100 employees. The few records representing large employers, (greater than 5000 employees) represented both public and private sector establishments. Tables 4 and 5 below show the distribution of establishments by total employees. The importance of this information is that a single address is associated with each unique record in these files. So, for instance, the employees of the establishment with 25K+ employees in Broward County all will be sent to a single location in an origin/destination trip table.

	Number of Establishments with > 25K Employees	Number of Establishments with 10 - 25K Employees	Number of Establishments with 5 - 10K Employees	Number of Establishments with 1 - 5K Employees	Number of Establishments with 100 - 1000 Employees	Number of Establishments with < 100 Employees
Florida						
Broward	1	0	3	13	160	12343
Miami-Dade	2	1	4	20	273	18435
Palm Beach	0	1	0	7	154	8972
District 4	0	0	0	3	45	2471
Rest of State	0	4	13	134	1673	69923

Table 4: Florida Establishments by Number of Employees.

	Number of Establishments with > 25K Employees	Number of Establishments with 10 - 25K Employees	Number of Establishments with 5 - 10K Employees	Number of Establishments with 1 - 5K Employees	Number of Establishments with 100 - 1000 Employees	Number of Establishments with < 100 Employees
Illinois	0	10	6	123	2152	140072

Table 5: Illinois Establishments by Number of Employees.

Representation of establishments occupying multiple locations may be achieved in the ES202 data sets by assigning unique unit numbers to each location. Tables 6 and 7 summarize the number of multi-unit establishments currently in the files sent to the states.⁴ The vast majority of information in the priority 2 files appears to be for single unit establishments (i.e. one address per establishment).

	Multi-Unit Businesses	Multi-Unit Employees
Broward	48	3326
Miami-Dade	52	14521
Palm Beach	35	1435
District 4	77	7372
Rest of State	937	390489

Table 6: Florida Multi-Unit Businesses.

	Multi-Unit Businesses	Multi-Unit Employees
Illinois	1300	493802

Table 7: Illinois Multi-Unit Businesses.

Results:

Table 8 summarizes the results of the look-up process by the states. An address is considered “returned” if the SEIN plus SEIN Unit are in the original file sent to the state. See the discussion of quality issues for additional details.

	Total Returned Addresses	Total Returned Single Unit Addresses	Total Returned Multi-Unit Addresses
Florida	11,932	11,382	550
Illinois	67345	45801	21544
	Total Returned Employment	Total Returned Single-Unit Employment	Total Returned Multi-Unit Employment
Florida	441,845	336,453	105,392
Illinois	702,634	673,629	29,005

Table 8: Look-up address summary.

Florida returned addresses for 10% of the establishments (representing 22% of the employment) sent to them for look-up and Illinois returned addresses for 47% of the establishments

⁴ These tables were generated using only the SEIN to represent an establishment. If the SEIN appears only one time, then the establishment is considered a single unit. Otherwise, it is a multi-unit. Unfortunately, the SEIN might appear in the master file more than once, but only have one priority 2 record because the address is missing for only one of the establishment’s locations.

(representing 40% of the employment). These numbers do not reflect the addresses returned to Westat that do not match to a SEINUNIT in the delivered files.

The “returned” addresses have not been geocoded so it is inappropriate at this time to comment on the quality of the address information. Westat geocoded the records with street addresses in the Palm Beach and Broward Co. files using GDT’s Dynamap 2002 database. Table 9 contains a summary of the batch match process. Over 90% of the look-up addresses matched to the block level.

	Palm Beach	Broward
Total Records	3452	3662
Records matched	3256	3420
Percentage matched	94%	93%

Table 9: Geocoding Results for Palm Beach and Broward Co.

Quality Issues:

MULTIPLE UNITS

It became apparent that some of the businesses that were identified by a single location actually should be represented by multiple locations for transportation purposes. Each state was sent a set of tables containing “Priority” establishments. These establishments were generally large employers represented by a single address.

Florida was sent 5 “Priority” datasets. One file contained 17 establishments (357,486 workers) for employers that have greater than 10,000 employees but only one unit. The other four files contained information on employers with 1,000 to 10,000 employees and only one unit. These were organized by type of organization (medical - 49 establishments, 112,178 workers, education - 45 establishments, 146,337 workers, public - 50 establishments, 129,708 workers, and, Private-miscellaneous - 181 establishments, 418,511 workers. The latter file contained several temp (staffing) agencies whose employees do not report to a single location. We recommended not focusing on temp agency addresses. However, there are several retail facilities where we needed store location and allocation of employees to each store.

Illinois was also sent 5 “priority” files. One file contained 7 establishments (157,323 workers) for employers that have greater than 10,000 employees but only one unit. The other four files contained information on employers with 1000 to 10,000 employees and only one unit. These were organized by type of organization (medical - 53 establishments, 120,943 workers, education - 53 establishments, 111,194 workers, public - 28 establishments, 64,730 workers, and Private-miscellaneous - 143 establishments, 278,355 workers. This last file contained several temp agencies as well.

Florida focused on the Education priority file. The 68 addresses turned into 2629 addresses for an estimated 257,219 workers at Florida public schools. However, the employment estimation procedure results in nearly 63,000 fewer employees than reported in the ES202 files. They also found nearly 70 addresses for Community Colleges.

Illinois returned all of their priority files. Table 10 summarizes their effort. Note that these files contain both establishments where the physical address geocoded to an acceptable level as well as those with no valid physical address. They converted 284 establishment records into more than 6464 physical locations.

PRIORITY	EDUCATION	MEDICAL	PRIV_MISC	PUBLIC	EMP10KPLUS	TOTALS
0	42	36	110	15	2	
2	11	17	33	13	5	
TOTAL	53	53	143	28	7	284
Multiple Addresses Returned	542	151	3489	1119	1163	6464
Employment	71662	55914	180317	53908	146783	508584

Table 10: Summary of Illinois priority files.

During the look-up process on the original files, both states returned multiple addresses for an establishment if they found multiple addresses. However, each group working on these files had a slightly different way of documenting this. For example, Table 11 shows that the MPOs working on Palm Beach and Broward Counties added new SEINUNITS whereas the others duplicated SEINUNITS if multiple addresses were found. Illinois created new SEINUNITS when multiple addresses were found. Illinois also attempted to allocate employment to the multiple units.

FLORIDA SUMMARY	Total Matches to P2	SEINUNITS in P2	% complete	Returned Addresses for unknown SEINUNITS
Palm Beach				
Total (1 to 1)	1701	9134	18.6%	537
Broward				
Total (1 to 1)	2686	12520	21.5%	976
Miami-Dade (1 to 1)	2017			
(1 to 2)	104			
(1 to 3)	10			
(1 to 4)	3			
(1 to 5)	4			
(1 to > 5)	5			
Total (1 to ?)	2143	18735	11.4%	0
District 4 (1 to 1)	200			
(1 to 2)	6			
(1 to 3)	2			
(1 to 4)	0			
(1 to 5)	0			
(1 to >5)	1			
Total (1 to ?)	209	2519	8.3%	0
Rest of State (1 to 1)	4778			
(1 to 2)	273			
(1 to 3)	57			
(1 to 4)	21			
(1 to 5)	21			
(1 to >5)	43			
Total (1 to ?)	5193	71750	7.2%	0

Table 11: Florida Address Summary for SEINUNITS

Information on the number of records with a street address only represents if the field is empty or filled (not necessarily geocodable).

For Illinois, Table 12 provides a summary of multi-units. Information on the percent of records investigated is also in this table. The last column indicates the number of unique SEINUNIT for which an address was returned but there is no corresponding SEINUNIT in the original database. This number DOES NOT include the 6180 additional addresses in the five priority files returned by Illinois.

ILLINOIS SUMMARY	Total Matches to original file	SEINUNITS in P2	% complete	Returned Addresses for unknown SEINUNITS
Illinois (1 to 1)	45836			
(1 to 2)	274			
(1 to 3)	16			
(1 to 4)	2			
(1 to 5)	1			
(> 5)	19			
Total (1 to ?)	46357	114363	41%	2884

Table 12: Completed investigation summary.

To help clarify the information presented above on single and multi-units, we looked at a single establishment in Florida. The web site for Firm AAA (who bought Firm BBB) has a store locator feature that generates 429 store addresses in the state of Florida. A search in the file containing all 2001 ES202 records yielded:

- 79 records, all with Unique IDs (SEINUNITS).
- Fourteen of these records have priority = 0 (no look-up) and 65 have priority = 2 (invalid physical address).
- There are 20 unique SEIN associated with the 79 records. However, one SEIN has 60 units accounting for 4743 of the 4815 employees.
- Seventeen of the 79 records have an address in the Code1 geocode field.
- Seven of these have the same street name and city as an address from the web site. All seven have different street numbers.

In the four SAS files returned by Florida there are 375 records associated with Firm AAAA and virtually all of the addresses associated with these records exactly match to one of the 429 on the web site. There are a few in these files that do not match to a web site address. There are 61 Unique IDs (SEINUNITS) associated with these 375 records and only 6 SEIN (one SEIN has 370 units). It appears that the files returned from Florida represent the locations of this establishment much better than the outgoing file. In other words, cleaning worked.

RINGERS

To verify the quality of the addresses sent back from the look-up process, Census included some establishments that already had a valid physical address. These establishments are called “Ringers”.

The information in the files returned from the states was compared to the information in the “Ringer files” provided by Census. In the Florida ringer file 205 of the records had no Code1 geocode address to which to compare the returned address. These 205 records were distributed as follows:

Broward = 30,
 Miami-Dade = 52,
 Rest of State = 95,

Palm Beach = 25,
District 4 = 3

The 2285 records with an address were distributed as follows:

Broward = 265,
Miami-Dade = 428,
Rest of State = 1345,
Palm Beach = 191,
District 4 = 55,
Statewide = 1

For Broward County, only one ringer was found and neither the address nor city matched the physical, mailing, or Code1 address in the ES202 file.

The Unique ID from the ringer file matched to 493 records in the Miami-Dade file. Only 79 of the matching records in the Miami file had street address information returned. Sixteen of these addresses were associated with 3 establishments. One establishment (Unique SEINUNIT) has 11 addresses associated with it, another has 3 addresses and another has 2 addresses where there is only one address in the original ES202 file. In all, 26 of the returned addresses match those in the Ringer file.

The Unique ID from the ringer file matched to 1474 records in the Rest of State file. 192 of the matching records had street address information. 105 of these have the same address as the corresponding record in the ringer file. (23 matched to Ringer records with blank addresses.)

Only 3 ringers were found in the Palm Beach files and only one of these had an address in the corresponding record of the ringer file. That address did not match.

In the District 4 file there are 58 ringers, 4 with a street address, but they match to blank addresses in the Ringer file.

One ringer is associated with an establishment that does business statewide but no address was returned.

The Illinois ringer file has 59 blank address fields. They are distributed as follows:

Cook = 29,
DuPage = 5,
OtherCty = 16,
Grundy = 1,
Kane = 2,
Kankakee = 1,
Lake = 2,
McHenry = 1,

Will = 2

The 1943 records with addresses are distributed as follows:

Cook = 853,
DeKalb = 9,
DuPage = 172,
OtherCty = 574,
Grundy = 5,
Kane = 63,
Kankakee = 16,
Kendall = 5,
Lake = 124,
McHenry = 46,
Will = 76

There are 432 ringers in Cook Co. and 71 match the geocoded physical address. McClean has addresses for 7 ringers and they all match. Kankakee has addresses for 11 ringers and they all match. The remaining areas have addresses for 190 ringers and 28 do not match.

In most non-matching cases, it appears that the returned address may be more accurate than the original Code1 address based on reverse phone look-ups and Internet searches performed by Westat.

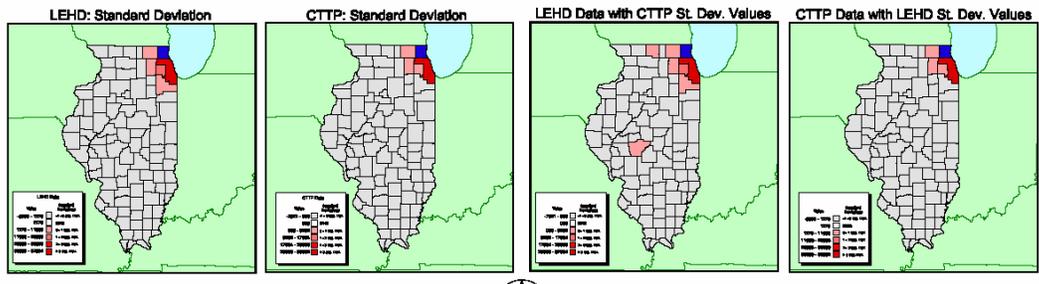
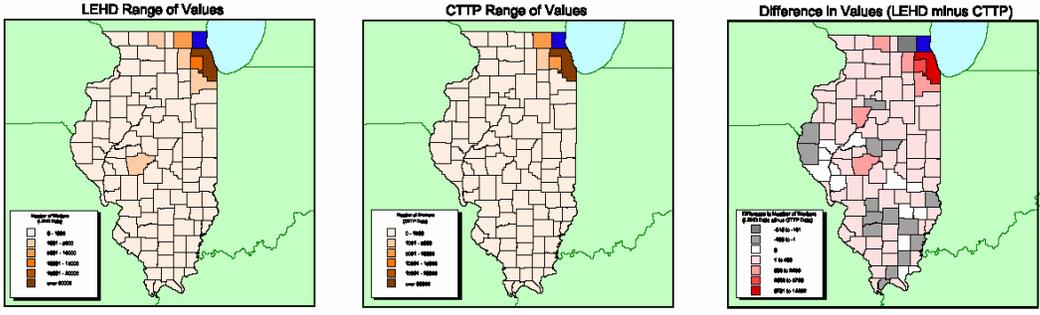
Comparison of LEHD data to CTPP data

We used the 1999 pilot to compare the work and residence locations derived from LEHD data with those derived from the CTPP for two counties: Lake County, IL and Miami/Dade Counties, FL. It is worth noting that we expect these comparisons to improve with 2001 data both because the full place of work multiple imputation procedure is applied and because of the incorporation of the place of work information by the states. It is also worth noting that these comparisons are at the county level only and that we expect the LEHD data will be more comparable to the CTPP and ACS data at lower level geographies such as tract and TAZ.

Briefly, both 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 are quite close. There are marked discrepancies – particularly for Florida – in the place of work information. While in Illinois the only marked discrepancy is the number of workers reported as working in Springfield – the capital – while working in Lake County, the reporting issues are much more troubling for FL – possibly reflecting the importance of personnel supply companies in the state.

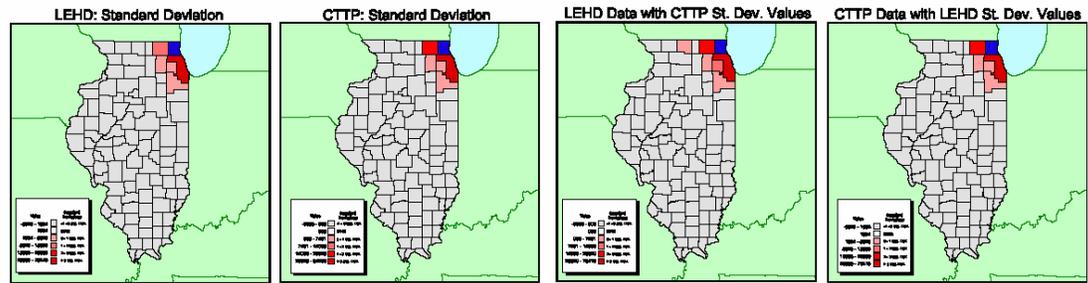
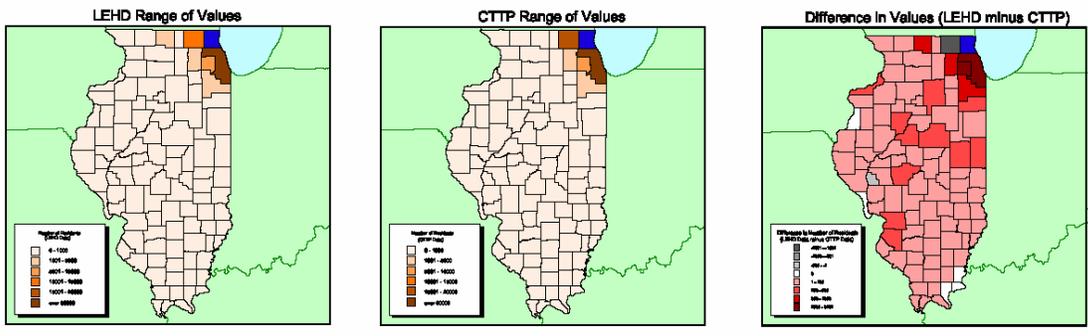
LEHD vs CTPP : COMPARISON OF DATA

Work Locations of Lake County Residents



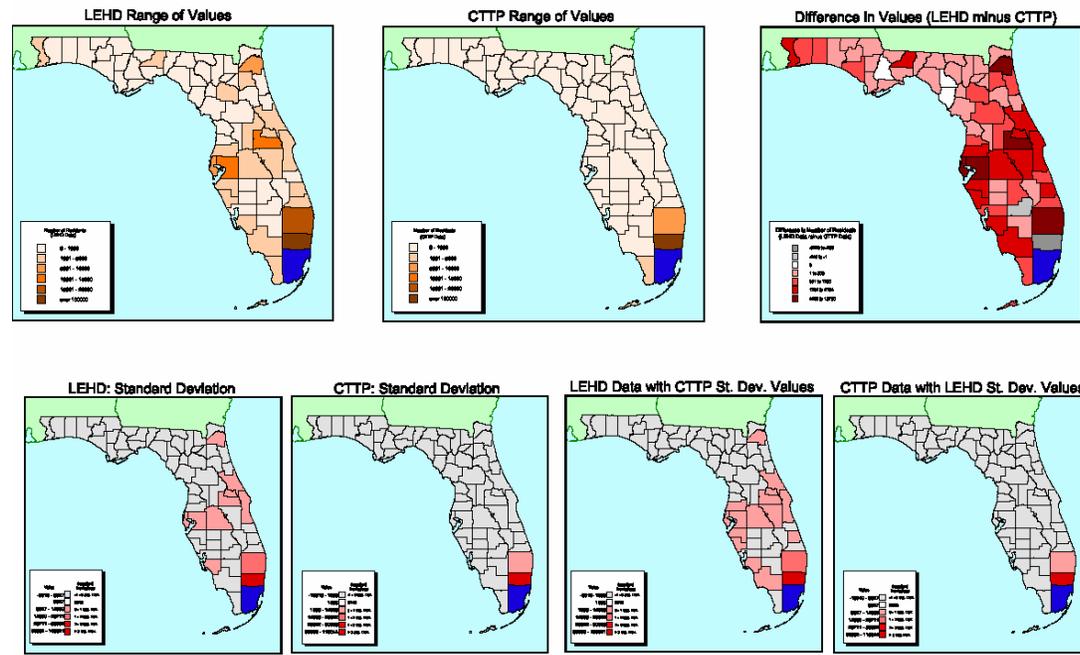
LEHD vs CTPP : COMPARISON OF DATA

Home Locations of Lake County Workers



LEHD vs CTPP : COMPARISON OF DATA

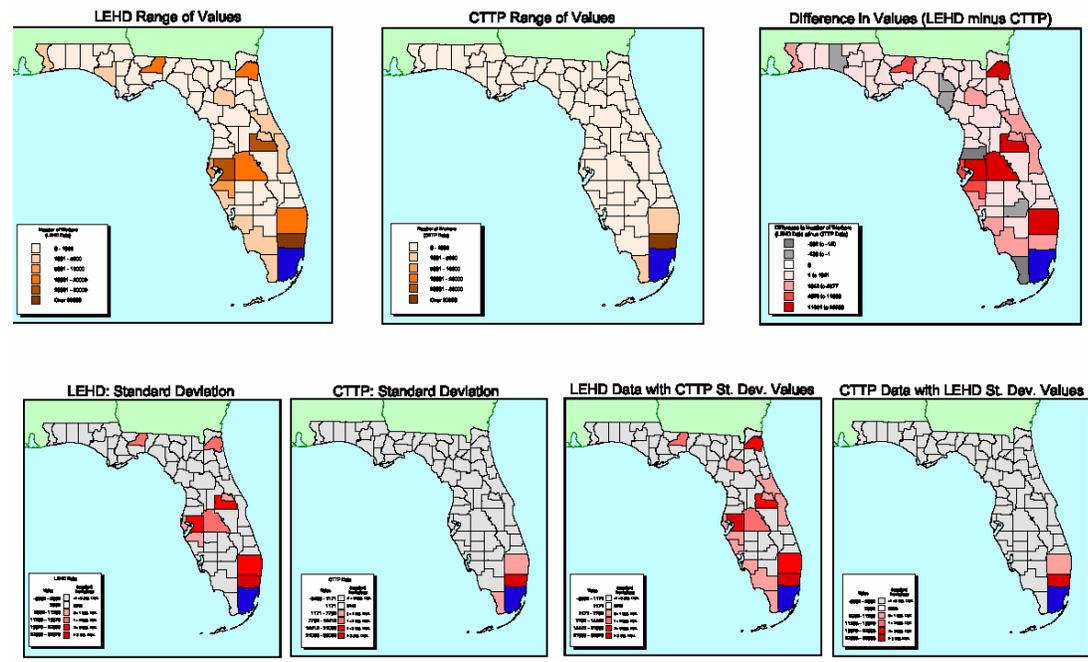
Home Locations of Miami-Dade County Workers



WESTAT
STATISTICAL CONSULTANTS
Page 24 of 28

LEHD vs CTPP : COMPARISON OF DATA

Work Locations of Miami-Dade County Residents



WESTAT
STATISTICAL CONSULTANTS
Page 24 of 28

Lessons Learned

A better approach would have been to ask the states to work on the entire ES202 files rather than prioritizing the records based on a geocodable physical address. Our assumption that an address was “good” just because it existed was not valid. As the program continues, validated addresses may be flagged and maintenance activities performed only on establishments not validated.

It would similarly have been useful to have made better use of SIC codes for look-up and validation. If establishments were sorted by type this would facilitate the look-up in a variety of sources and prevent look-ups of establishments whose workers probably do not report to a single location each day.

Improving the multi-unit breakouts would be easier if the focus were first on files where additional administrative or Internet resources were available. The most successful breakouts were with the education file that had extensive administrative data available from Florida DOE.

The MPOs should be brought into the look-up process as soon as possible and have regularly scheduled conference calls with each participant agency to check progress, answer questions, provide technical assistance and encouragement.

It would be useful to ask states to incorporate multi-units into their systems and maintain these as multi-units. If this part of the program were to become ongoing, the eventual goal would be to have each storefront, office space, or location occupied by an employer to have an address and valid number of employees working at that address.

It would also be useful for the Census Bureau to routinely incorporate non-Title 13 geocoding in files that are returned to the states.

Next Steps

A key component is to expand the LEHD program to all 50 states, and territories. This would permit the analysis of transportation issues that span state borders.

The two states involved in this pilot – Florida and Illinois - could spearhead a broader approach in multiple states. In particular, their experiences could be used to develop an ongoing cooperative relationship with states – particularly the local transportation agencies to improve the ES202 (establishment file).

The Census Bureau could evaluate the value added from matching the ES202 data to the Census Business Register and the Census of Governments, to improve the quality of the physical addresses and the multi-unit breakout information.

Finally, the Census Bureau could leverage both the LEHD data and ACS data by integrating the two data sources to model commuting modes and the time of day of the commute.

Appendix 1: Palm Beach Report

February 7, 2003

LEHD Review - Palm Beach County

The attached database file represents my review of the LEHD files provided. The file contains the 'control number', sub number with 3 data fields for joining to the master file.

COMMENTS

- | | |
|--------------------|-------------------------------------------------------------|
| 1. IU | InfoUSA Match |
| 2. OCCLC | Local occupation license file match |
| 3. CANNOT VERIFY | Could not verify this entry |
| 4. CENTRAL OFFICE. | This entry is most probably not a central business location |
| 5. INTERNET | Found on Internet |

FNLSTREET Reviewed street address

FINALZIP Reviewed zip code

Total time on project - **20 hours**

All comments of 'CANNOT..' and 'CENTAL...' were individually reviewed.

All FNLSTREET entries not commented were individually reviewed and entered. Each was compared against the addresses in the provided files and in many cases these were sufficient.

Say *1(one) hour* for the InfoUSA join and population of table.

Say *2(two) hours* for the occupation license file. The master file was reviewed to identify duplicate names so they would not be used in the join and population of the LEHD file.

Recommendations

The level of effort to make this file 'clean' is only the result of not having a fairly reliable situs address in the original file. Considering this is not for public use, this should be resolved through file maintenance by the appropriate department. Obviously, there will not be a complete match during the geocoding effort but if situs addresses were available, the value of only those entries that are defined would probably meet any demand of value for the intended use.

Paul Larsen
c/o Palm Beach MPO
160 Australian Ave - Ste 201
West Palm Beach, FL 33406

Appendix 2: FLORIDA SUMMARY REPORT

LEHD

Broward and Palm Beach County Business Address Search Summary (for business with 2 to 24 employees)

Step 1: Use InfoUSA database to geocode the second delivery of employer CD. On the second CD, Broward and Palm Beach County have 17,069 and 12,287 records, respectively. Based on the InfoUSA database, 2,320 records have been matched for Broward County and 1,656 records have been matched for Palm Beach County.

Step 2: Separate two counties' matched records and unmatched records. Then, focus on records unmatched by the InfoUAS of these two counties. Try to find out the locations of the businesses without both mailing addresses and physical addresses. The records that have addresses in the first delivery were ignored.

Step 3: The businesses that have employee numbers equal to and great than 25 were given to Broward (? records) and Palm Beach (? records) Counties. District 4 Planning Office only works on these two counties' records with number of employees between 2 and 24, inclusively. For businesses that have only one employee, we assume that the owner works at home so no need to find out the business address. Consequently, within this framework, Broward County has a total of 1,483 businesses that need to have the physical addresses searched. Among these businesses, 728 records have no physical addresses, and 754 business addresses are found. Palm Beach County has a total of 1,125 businesses needing physical addresses search. Among them, 539 records have no physical addresses, and 586 business addresses are found.

Step 4: The method to perform the address search is described in detail below.

- All companies for Broward and Palm Beach Counties were searched by phone number and company name using *search bug.com* and *yahoo yellow pages*. *Search Bug.com* allows a search by company name and/or phone number. If a company name was not located, the phone number was searched. When there was no listing for the company by phone number, the phone number provided with employer CD was then called.
- When a number was called and the phone number was no longer in service or the number was recycled, (meaning that the number is now for a private resident or a different company,) “**No listing for company**” was placed in the PF_City field. If the companies' headquarter number was given, the company location in the designated county was used. If there was no location in Florida, “**no fla - local**” was placed in the Access database. Some companies did not have offices in two counties, but instead, employees worked from home. In this case, it was indicated as “**work from home.**”
- When a number was called and a voice mail was reached, in that instance, “**voice mail**” was placed in the database. These numbers will be called again in order to locate the

correct addresses. When a phone number was not provided and the company could not be located by name, **“no phone # provided”** was placed in the PF_City field.

- When the provided phone number was no longer in service and there were several locations in the county, **“wrong #”** was used (i.e., multiple site Chain Store). With out the phone number, the correct store location could not be found. There were several companies that would not provide their information, then, **“can not help me”** was used.
- The detailed unmatched reasons are listed in the below two tables.

BROWARD COUNTY

	573 companies	Were not listed by name or phone # (phone # available)	“No listing for company”
1	67 companies	Were not listed by name and a phone # was not provided	“No phone # provided”
2	51 companies	Were not located in Florida	“No fla local”
3	22 companies	Had the wrong phone #, not able to locate their address in the county	“Wrong #”
4	8 companies	Had employees that worked from their homes	“Work from home”
5	5 companies	Could not be located through a search and employees would not give out location	“Can not help me”
6	2 companies	Were no longer in business	“No longer in business”

PALM BEACH COUNTY

1	357 companies	Had less then 2 employees	
2	399 companies	Were not listed by name or phone # (phone # available)	“No listing for company”
3	54 companies	Were not located in Florida	“No fla local”
4	53 companies	Were not listed by name and a phone # was not provided	“No phone # provided”
5	20 companies	Had the wrong phone #, not able to locate their address in the county	“Wrong #”
6	7 companies	Could not be located through a search and employees would not give out location	“Can not help me”

7	2 companies	Did not provide a store # and was not able to locate the correct store	“Which local”
8	2 companies	Had employees that worked from their home	“Work from home”
9	1 company	Were no longer in business	“No longer in business”

Approximately, a total of 10 person-working days were spent on these address searching.

Appendix 3: REPORT ON FLORIDA PRIORITY FILES

Florida Experience and Recommendations on Improving Origin and Destination Data by Traffic Analysis Zones (TAZ)

Background

The Florida Agency for Workforce Innovation (AWI), Labor Market Statistics (LMS) entered into a Memorandum of Agreement (MOA) with the Demographic Surveys Division, US Census Bureau to improve the quality of physical address information on state ES-202 Enhanced Quarterly Unemployment Insurance (UI) files. The purpose of the MOA was to test, develop, and recommend procedures for obtaining better physical address for businesses so that they could be geocoded.

The files were split out for public, private, medical, large establishments, and education employers in Florida. Another file contained just duplicate record files for Miami-Dade County. From the outset it was decided to eliminate the file with just large employers and combine this with the other files based on how the files would be refined. All government files were combined, as were all education and private establishments. Employers in the employee leasing and staffing industry (SIC 736) were eliminated due to the impossibility of achieving breakouts. The five remaining files were then analyzed separately to identify the problems associated with each file and how it would be resolved.

Methodology and Results

The first file to be resolved was medical since these were large hospitals in Florida with single employee parking garages for outpatient, hospital, and doctor's offices. Despite different departments, employees worked in the same physical locale. Consequently, there was no need to provide other physical addresses and employment.

The file for Miami containing duplicate addresses from ES-202 and InfoUSA was sent to the South Florida Regional Planning Council for technical assistance in refinement. The Regional Planning Council entered into an agreement with AWI to have access to confidential ES-202 individual firm records.

The Regional Planning Council was unable to allocate the resources needed to assist LMS before the deadline. Staff in Labor Market Statistics completed the file. Most of these duplicate records had inconsistent address fields between UI and InfoUSA and these were refined using Internet resources and phone calls. Some of the multi-units could not provide breakouts.

The education file was divided into community colleges and Florida school district offices. The Florida Department of Education was contacted for information on individual school addresses. These were determined to be available for individual schools under each school district. From the same file, each school had the number of teachers on the payroll and administrative and support staffs were estimated from the staffing information from the Occupational Employment

Statistics Survey. Several schools were contacted to verify if this procedure worked to secure an all employment figure. The procedure was validated. Employment at school district headquarters was estimated from the total UI figure minus the schools. School districts generally do not provide the number of employees at district headquarters.

The Florida Department of Education (DOE), Division of Community colleges could provide addresses for each of the state's community colleges. However, they were unable to provide employment for each community college campus. Permission was granted for LMS to call each community college. These calls revealed that most of the community colleges could not provide breakouts of employment due to the electronic handling of payrolls. Many adjunct professors teach at multiple campuses. The Director, Division of Community Colleges verified this problem and the impossibility of breaking out community college employment.

Little success was achieved with public employers, again because of the lack of administrative records to break out small offices such as tax collection or annex courthouses. Payrolls are handled electronically from the city hall or county court house for all city or county offices.

The files with the biggest problems and the least success in securing place-of-work addresses and employment breakouts were in the private miscellaneous file. These businesses had already refused to break out employment in the Bureau of Labor Statistics Multiple Worksite Report (MWR) survey for the ES-202 program. The majority of these employers could not provide breakouts of their employment.

Recommendations

Files where additional administrative or Internet resources were available could be more easily refined than those with just UI or InfoUSA as resources. The most successful breakouts were with the education file that had extensive administrative data available from Florida DOE.

Due to the difficulty of trying to achieve breakouts on the private and public (community college, county, and city governments) files, the only known successful approach would be to identify some other types of administrative records that could be used to break out locales. This leaves a problem with reporting employment. Some success was achieved in the past with county tax files. Clearly these breakouts would need to be done by county planning departments or some other local planning entity. A more comprehensive and coordinated effort with Regional Planning Councils and the Florida Department of Transportation would assist in this effort.

The use of administrative records worked for Florida's Public Schools only because employment was included in administrative records along with physical addresses. These records were matched to link employment with the address. Community colleges and municipal and county governments in Florida are generally not capable of providing breakouts of employment due to problems with how payrolls are handled.

