

What's new in LEHD Snapshot S2014

Lars Vilhuber

Center for Economic Studies, US Census Bureau
Labor Dynamics Institute, Cornell University

[lars.vilhuber@{cornell.edu,census.gov}](mailto:lars.vilhuber@cornell.edu)

September 14, 2017

Disclaimer

Disclaimer

This presentation reports the results of research and analysis undertaken by 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 document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed herein are attributable only to the author(s) and do not represent the views of the U.S. Census Bureau, its program sponsors, Cornell University, or other data providers.

The LEHD Infrastructure Files (“Snapshot”)

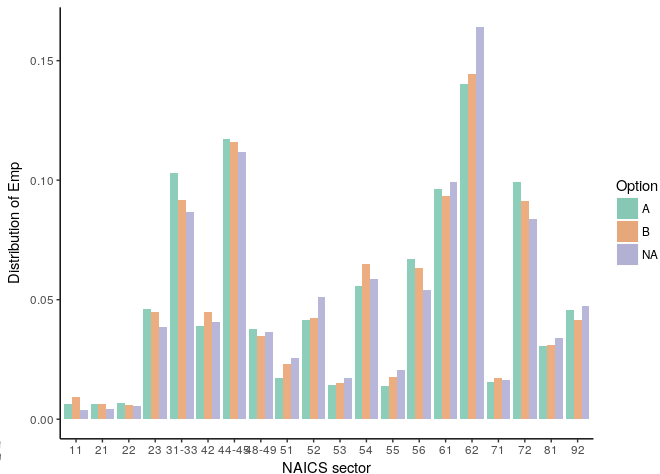
- ▶ Big-picture overview in Abowd et al. (2006, 2009)
- ▶ Defining outputs: Quarterly Workforce Indicators (QWI) and to some extent the LEHD Origin-Destination Employment Statistics (LODES) data
- ▶ Files are pulled from LEHD production archives, several research-related improvements are made to the files, fixing minor data inconsistencies or updating documentation.
- ▶ Do not contain any information related to the disclosure limitation measures used in the QWI

Availability by MOU

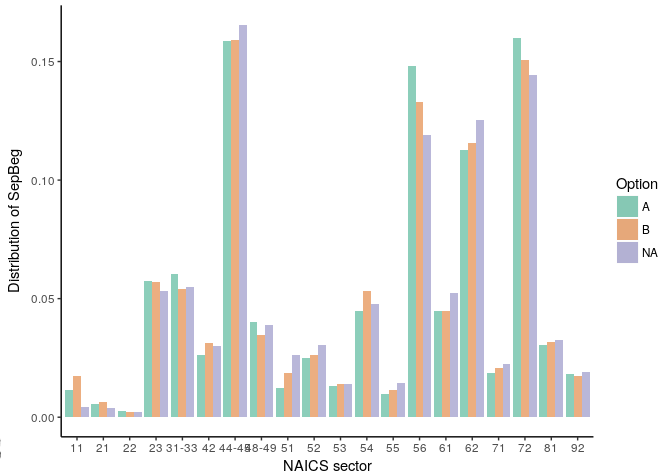
Of the 51 states and DC, as of 2017-09-08, 12 have chosen Option A:

Option		Freq
A	Data available subject to CB approval and pooling	12
B	Data available subject to state approval	32
	of which frequently approve projects:	13
NA	Other	7

By employment



By separations



Representativeness

Is the sample of states representative = state's choice of Option A (M) correlated with any observable characteristic (Y):

$$p(M|Y, \theta, \psi) = p(M|\psi)$$

where θ are parameters associated with the data generating process, and ψ are parameters associated with the choice of Option A. If not, then missing states are “Missing Completely at Random (MCAR)”.

Generalisation

Replace “Option A” with “is part of my research study”

MCAR by employment

	Option
Constant	0.888*** (0.094)
Emp	0.295*** (0.068)
Akaike Inf. Crit.	1,082.799

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

MCAR by employment (b)

	Option
Constant	0.198*** (0.076)
Emp	-0.124*** (0.035)
Akaike Inf. Crit.	1,400.767

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

MCAR by separations

	Option
Constant	0.979*** (0.089)
SepBeg	0.204*** (0.058)
Akaike Inf. Crit.	1,093.073

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

MCAR by both

	Option
Constant	0.884*** (0.094)
Emp	0.403*** (0.130)
SepBeg	−0.104 (0.101)
Akaike Inf. Crit.	1,083.793

Notes:

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Representativeness

Approximately...

- ▶ The snapshot is *approximately* representative (eye-ball test),
- ▶ ... but that statement does not hold up to an econometric test
- ▶ To apply this particular criterion to your specific paper, see **github.com/larsvilhuber/snapshot-availability**

The following components have changes

- ▶ **Employment History Files (EHF)**: earnings at all firms for workers
- ▶ **Employer Characteristics File (ECF)**: time-varying characteristics for firms, including industry and QCEW employment
- ▶ **Geocoded Address List (GAL)**: geo-coded addresses of firms from a variety of sources
- ▶ **Office of Personnel Management (OPM)**: earnings records for federal workers
- ▶ **QWI(-SEINUNIT)**: pre-computed job market indicators (separations, job creation, etc.) at the establishment level, by worker demographics

EHF: New National Indicator of Employment

Does a worker have covered employment in **any** state?

Construction

For each PIK i , in each EHF state file s , identify the quarters q in which at least \$1 was earned in any job j :

$$\forall i, \forall q, \forall s : pos_earn_{isq} = 1 \text{ if } \left(\sum_{\substack{j: J(i,q)=1 \\ \wedge J(s,q)=1}} earn_{ijq} \right) \geq 1$$

where $J(i, q)$ indicates that i worked for j in q and $J(s, q)$ indicates that j is located in s in q .
Count the **number** of states (not jobs):

$$\forall i, \forall q : pos_earn_{iq} = \sum_s pos_earn_{isq}$$

EHF: New National Indicator of Employment

The file is not subject to state-specific restrictions, and is automatically added to any project requesting LEHD data.

EHF availability metadata

Temporal availability of states (`ehf_all_availability`), so that users can control for the at-risk states at each point in time

Restructured ECF

Compatibility

- ▶ Not completely backward compatible: some variables changed names / no longer produced
- ▶ **No changes** have been made to T26 variables → no change in permissions required

NAICS and SIC

- ▶ NAICS2012 codings are now available, in addition to other NAICS vintages
- ▶ SIC1987-based codings still available, consistently named

GAL restructured

- ▶ Completely restructured in production system
- ▶ An attempt was made to maintain backward compatibility, but **not all changes** could be reasonably mapped.
- ▶ **Crosswalks** to all input sources have been **re-introduced**.

Office of Personnel Management (OPM)

- ▶ OPM data on Federal workers have been collected under a **single directory**. RDC users should be able to access these files by requesting a “OPM” dataset.
- ▶ Access to the OPM data **do not require state** permissions.

Quarterly Workforce Indicators (QWI) SEINUNIT

Renamed and dropped variables

- ▶ In line with changed publication of QWI (lehd.ces.census.gov/doc/Memo_changes_to_QWI.pdf), several variables are no longer present, or have been renamed.
- ▶ In general, names are consistent with the "Alternate names" in (LEHD Schema V4.0.1)
- ▶ Coding for race and ethnicity has changed, this affects the **naming** of variables.

QWI SEINUNIT

The following variables are no longer available:

- ▶ All variables related to *changes in total earnings* (starting with dW)
- ▶ All variables that are rates (ending in R)
- ▶ All variables related to periods of non-employment preceding or following a transition (starting with N)
- ▶ Certain average earnings variables (WCA, WCS, WA, WS)
- ▶ Certain variables relating to continuous quarter hiring (CH, CR)
- ▶ The variable FS_Nx (Full-quarter separations in the **next quarter**)
- ▶ The variable W2 (average earnings for **end-of-quarter** employment), replaced by W2B (average earnings for **beginning-of-quarter** employment)

When is it available to researchers

Integrated Research Environment (IRE)

The S2014 Snapshot will become available when the IRE is made accessible to FSRDC researchers.

Introduction
S2014
What's new
Availability

The End

References



Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon D. Woodcock. 2006. *The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators*. Technical report TP-2006-01. U.S. Census Bureau, LEHD and Cornell University. <http://econpapers.repec.org/paper/centpaper/2006-01.htm>.



———. 2009. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts*, edited by Timothy Dunne, J. Brad Jensen, and Mark J. Roberts. University of Chicago Press. <http://www.nber.org/chapters/c0485.pdf>.



McKinney, Kevin, and Lars Vilhuber. 2011a. *LEHD Data Documentation LEHD-OVERVIEW-S2008-rev1*. Working Papers 11-43. Center for Economic Studies, U.S. Census Bureau, December. <http://ideas.repec.org/p/cen/wpaper/11-43.html>.



———. 2011b. *LEHD Infrastructure Files in the Census RDC: Overview of S2004 Snapshot*. Working Papers 11-13. Center for Economic Studies, U.S. Census Bureau, April. <http://ideas.repec.org/p/cen/wpaper/11-13.html>.



Vilhuber, Lars, and Kevin McKinney. 2014. *LEHD Infrastructure files in the Census RDC - Overview*. Working Papers 14-26. Center for Economic Studies, U.S. Census Bureau, June. <https://ideas.repec.org/p/cen/wpaper/14-26.html>.