

This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Producer Dynamics: New Evidence from Micro Data

Volume Author/Editor: Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 978-0-226-17256-9

Volume URL: <http://www.nber.org/books/dunn05-1>

Conference Date: April 8-9, 2005

Publication Date: January 2009

Chapter Title: Comment on "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators"

Chapter Author: Katharine G. Abraham

Chapter URL: <http://www.nber.org/chapters/c0486>

Chapter pages in book: (230 - 234)

- Abowd, J. M. and L. Vilhuber. 2005. The sensitivity of economic statistics to coding errors in personal identifiers. *Journal of Business and Economic Statistics* 23 (2): 133–52.
- Benedetto, G., J. Haltiwanger, J. Lane, and K. McKinney. 2007. Using worker flows in the analysis of the firm. *Journal of Business and Economic Statistics* 25 (3): 299–313.
- Bureau of Labor Statistics. 1997a. *BLS Handbook of Methods*. U.S. Bureau of Labor Statistics, Division of Information Services, Washington D.C. Available at <http://www.bls.gov/opub/hom/>
- . 1997b. Quality improvement project: Unemployment insurance wage records. Report of the U.S. Department of Labor.
- Chiang, H., K. Sandusky, and L. Vilhuber. 2005. Longitudinal Employer-Household Dynamics (LEHD) Business Register Bridge technical documentation. *Internal Document IP-LEHD-BRB*. LEHD, U.S. Census Bureau.
- Davis, S. J., J. C. Haltiwanger, and S. Schuh. 1996. *Job creation and destruction*. Cambridge, MA: The MIT Press.
- Longitudinal Employer-Household Dynamics Program. 2002. The Longitudinal Employer-Household Dynamics program: Employment Dynamics Estimates Project version 2.2 and 2.3. *Technical Paper TP-2002-05-rev1*. LEHD, U.S. Census Bureau.
- Stephens, B. 2006. *Firms, wage dispersion, and compensation policy: Assessment and implications*. Ph.D. diss. University of Maryland, College Park, Maryland.
- Stevens, D. W. 2007. Employment that is not covered by state unemployment insurance laws. *Technical Paper TP-2007-04*. LEHD, U.S. Census Bureau.

## Comment Katharine G. Abraham

This chapter describes in some considerable detail the sources and methods used to construct the data files that underlie the new Quarterly Workforce Indicators (QWI) produced by the U.S. Census Bureau. This innovative program draws on a wide variety of data sources to produce county-level estimates of earnings, employment, and job flows, disaggregated by industry, age of worker, and sex of worker. The resulting estimates already have proven to be of considerable interest to local planners and policymakers, and it is easy to imagine additional uses for them. The chapter should be a valuable resource for users of the QWI data as well as for researchers who may be interested in working with the underlying data files.

Unavoidably, given the ambitious nature of the exercise undertaken and the limitations of the underlying source data, development of the QWI has confronted a variety of data problems. The QWI files draw heavily on administrative records—including unemployment insurance (UI) wage

Katharine G. Abraham is a professor in the Joint Program in Survey Methodology and a faculty associate of the Maryland Population Research Center at the University of Maryland, and a research associate of the National Bureau of Economic Research.

records, employer reports to state employment security agencies, and the Census Bureau's Person Characteristics File based on the Social Security Administration's Numident file—which were not developed for statistical purposes. Other information is drawn from large national surveys that have better statistical properties but cover only a fraction of the population. Much of the chapter is devoted to explaining the methods currently used to address the various shortcomings of the underlying source data, as well as improvements in those methods planned for the future. My comments review briefly some key issues that the QWI developers have had to confront.

- *Miscoding of individual identifiers.* If not corrected, miscoding of individual identifiers will lead to overstatement of worker flows, misrepresentation of workers' earnings trajectories, and misstatement of the earnings of both departing and newly hired workers. A 1997 study of UI wage records conducted by the Bureau of Labor Statistics found that approximately 7.8 percent of individual Social Security numbers were miscoded (U.S. Bureau of Labor Statistics 1997). Abowd and Vilhuber (2005) describe a clever automated method for identifying and correcting miscodes that may occur in the middle of an ongoing spell of employment, but this method cannot capture coding mistakes that are caught by reporters and permanently corrected, or coding mistakes that are never caught. By design, the Abowd and Vilhuber procedure is conservative, producing recodes for only about 0.5 percent of wage records. While they are somewhat dated, the larger figures from the BLS study suggest that there may be a substantial amount of miscoding in individual identifiers that the Abowd and Vilhuber procedure does not capture. Further research will be needed to determine the severity of individual identifier miscoding, and what it implies for various potential uses of the QWI and associated data files.
- *Failure to identify continuing firms or establishments with new identification numbers.* Similar to the problems associated with miscoding of individual identifiers, treating continuing establishments as new businesses leads to overstatement of business births and business deaths, as well as to overstatement of worker flows. This is perhaps the most-studied of all of the various potential problems with the QWI source data, and the techniques employed to identify establishment matches in business register data have improved a great deal over the past ten years. A clever recent innovation pioneered in the course of developing the QWI is the use of information on flows of groups of workers across establishments to identify cases in which a firm that appears to be a new birth is really a reincarnation of an old firm. While there undoubtedly are remaining cases in which continuing businesses are not

identified as such, this has to be a less serious problem than it would have been even a few years ago.

- *Missing information on individual characteristics.* In the QWI, missing information on individual characteristics is filled in using multiple imputation techniques. Information on individuals' age and gender is derived from Social Security records and is missing for just 3 percent of QWI records. Place of residence is missing for about 10 percent of records. The only individual-level information on education presently available for use in building the QWI files is that derived from the Survey of Income and Program Participation (SIPP) and Current Population Survey (CPS), meaning that education is missing and must be imputed for most records. This is done based on the relationship of education to age, earnings and industry in the 1990 Census. The very high rates of imputation for education cannot help but make users of these data uneasy. The planned incorporation of direct information on education for the approximately one-sixth of the population that completed the 2000 Census Long Form will be a positive step, but the share of people for whom education must be imputed will remain large.
- *Missing information on employer characteristics.* Employer-provided information contained in the business register files is used to assign NAICS codes and a geographic location to establishments, as well as to characterize the structure of the firms to which these establishments belong. Though specific percentages are not cited, a significant number of imputations must be performed to produce a complete data file (see Konigsberg et al. 2005, for a discussion of allocations and imputations in the Quarterly Census of Employment and Wages based on the same employer characteristic source data as the QWI). The best imputations likely are those that can be based on records for the same establishment from other time periods; such information, however, is not always available. As with the data for individuals, the use of imputed information on employer characteristics may be a problem for analytical uses of the data.
- *Missing information on the specific establishment in which each worker is employed.* When a firm consists of just one physical establishment, there is no difficulty in determining where a person employed by that firm works; in cases where the firm has more than one establishment, however, the assignment of individual workers to specific establishments generally is not reported. Only in Minnesota do the UI wage records indicate which establishment of a multiple-establishment firm employs which workers. As described in the chapter, the data for Minnesota are used to develop a model for probabilistically assigning workers to specific worksites within their firm that is then applied to the information available for other states. Whether a model fit using

Minnesota data can reasonably be applied to other locations is, of course, very much an open question. One of the most intriguing uses of the QWI data files is to analyze the geography of economic development, looking, for example, at where people live, where they work, and the patterns of travel between those locations. Errors in the assignment of workers to establishments could be especially problematic for this sort of analysis.

In addition to these data quality issues, the chapter also notes current limitations in the scope of the QWI data set. Two in particular seem important. First, it is not presently possible to track workers who move from one state to another. Second, the self-employed are presently excluded from the QWI universe. Depending on the question one was interested in answering, both of these exclusions could be substantively important. If, for example, significant numbers of displaced workers move into self-employment, using the QWI data to study the earnings consequences of job loss could produce misleading conclusions. The chapter indicates that work is underway to address these current limitations of the QWI.

A final point to note is that noise is added to the QWI records to protect the confidentiality of the underlying information. The designers of the process used to fuzz the QWI data pay attention to preserving their statistical properties, and the chapter suggests that the analytic validity of the files should not be adversely affected. This can be asserted confidently, however, only with respect to the examination of relationships that were anticipated in the design of the fuzzing process.

The preceding comments are in no way intended to be critical of the authors or to disparage the work that has been done to produce the Quarterly Workforce Indicators. As a practical matter, there is no real alternative to the use of administrative statistics to produce local labor market information at the level of detail contained in the QWI. Further, though they are sometimes discussed in a way that suggests they can be taken as truth, survey data also suffer from a variety of sampling and nonsampling errors. These are seldom as well documented as the potential errors in the QWI described in the chapter, but that does not mean they do not exist.

Still, it is important to recognize and remember that a good deal of the information that underlies the Quarterly Workforce Indicators is imputed rather than measured directly. In some cases, this will not matter very much; in other cases, the use of imputed data could lead to results that are misleading. Given the complexity of the process used to construct the indicators, it is rather difficult to know what degree of confidence to place in the picture they paint. Documenting the methods used to construct the data is an important first step and one the authors are to be commended for having taken. Further work will be required to develop a fuller under-

standing of the quality properties of the QWI estimates and data files, and of their suitability for different analytic purposes.

### References

- Abowd, J. M., and L. Vilhuber. 2005. The sensitivity of economic statistics to coding errors in personal identifiers. *Journal of Business and Economic Statistics* 23 (2): 133–52.
- Konigsberg, S., M. Piazza, D. Talon, and R. Clayton. 2005. Quarterly Census of Employment and Wages (QCEW) Business Register Metrics. Paper presented at the Joint Statistical Meetings, August. Minneapolis, Minnesota.
- U.S. Bureau of Labor Statistics. 1997. *Quality improvement project: Unemployment insurance wage records*. Unpublished report, Washington, D.C.