LONGITUDINAL ESTABLISHMENT MICRODATA AT THE BUREAU OF LABOR STATISTICS: DEVELOPMENT, USES, AND ACCESS

James R. Spletzer, Bureau of Labor Statistics
2 Massachusetts Avenue NE, Suite 4945, Washington DC 20212

Key Words: Longitudinal establishment microdata, job creation and job destruction, net employment growth.

1. Introduction

The Bureau of Labor Statistics (BLS) is currently constructing a longitudinal database that will contain quarterly employment and wage data for virtually all business establishments in the United States. This longitudinal database will enable us to track changes in employment and wages not only at the macro level, but also at the micro level of the establishment. This database, referred to as the Longitudinal Database or LDB for short, will be used to generate high quality, high frequency, timely and historically consistent information regarding not only job creation and job destruction, but also the life cycle of establishments. The purpose of this paper is to describe this database and to present some research results that will highlight the uses of this database.

2. Motivation

Although the cross-sectional or "snap-shot" employment statistics that are published by the government statistical agencies are invaluable for policy-makers, researchers, and the business community, these data are unable to completely describe changes in employment and/or the number of establishments over time. Comparing aggregate employment levels at two points in time only tells us about the net change in employment, and does not inform us with regard to how many establishments are either expanding or contracting, nor by how much these establishments are either expanding or contracting.

Job creation is defined as the employment growth contributed by establishments that expand or start up, and job destruction is defined as the employment decline resulting from establishments that contract or shut down. The sum of job creation and job destruction is the net change in employment. This decomposition of net employment change is important because it illustrates the underlying level of volatility in the labor market. This can easily be seen by an example.

Assume that payroll employment in December 1996 is 120,659,000 jobs, and that payroll employment in March 1997 is 121,344,000 jobs. Net employment growth is 685,000 jobs during the quarter. This net employment growth of 685,000 jobs is consistent with many scenarios, including either of the following three: 1) 685,000 jobs created and 0 jobs destroyed, 2) 10,022,620 jobs created and 9,337,620 jobs destroyed, 3) 121,344,000 jobs created and 120,659,000 jobs destroyed. Scenario #1 illustrates a labor market where no employer decreased the size of his establishment. Scenario #3 illustrates a labor market where all establishments in the previous quarter shut down and all establishments in the current quarter started up. The true underlying labor market is, of course, somewhere in between these two extreme cases. Scenario #2 illustrates an intermediate case, where the job creation rates and job destruction rates are roughly 8 percent.

Since a simple comparison of two cross-sectional aggregate employment levels does not provide any information regarding the underlying level of volatility in the labor market, it is clear that longitudinal microdata at the establishment level is required to decompose net employment change into its components of job creation and job destruction. This decomposition is one of the primary purposes of the LDB currently being developed.

Much of what we know about job creation and job destruction comes from research conducted at the Center for Economic Studies at the U.S. Bureau of the Census using the Longitudinal Research Database (the LRD); see, in particular, Davis, Haltiwanger and Schuh (1996). Despite all that has been learned about the labor market from the LRD, the conclusions that can be drawn from these data are somewhat limited since they only cover the manufacturing sector of the economy. Manufacturing in 1995 accounted for only 19 percent of private sector employment. Recent work by Anderson and Meyer (1994), Lane, Stevens, and Burgess (1996), and Spletzer (1995) has illustrated how job creation and job destruction in manufacturing may not be representative of the entire economy. Since the longitudinal database currently under construction at the BLS will essentially be a quarterly census of establishments and will thus encompass all industries, the job creation and job destruction statistics derived

from the LDB have the potential to be among the most important economic indicators published by the statistical agencies of the U.S. government.

3. Data Sources and Definitions

The source of the establishment microdata used for the LDB will be the ES-202 program, which is a cooperative endeavor of BLS and the states. All employers subject to state unemployment insurance (UI) laws are required to submit quarterly contribution reports detailing their monthly employment and quarterly wages to the State Employment Security Agencies. After the microdata are edited and, if necessary, corrected by the State Labor Market Information staff, the states submit the data to the Bureau of Labor Statistics as part of the Covered Employment and Wages program (ES-202). The data gathered in the ES-202 program are a comprehensive and accurate source of employment and wages, and provide a virtual census (98 percent) of employees on nonfarm payrolls. According to *Employment and Wages*, employers in private industry in 1996 provided state employment security agencies with quarterly UI tax reports for an average of 99.3 million wage and salary workers in approximately 6.9 million business establishments. For more information on the ES-202 program, see U.S. Bureau of Labor Statistics (1997) and Farmer and Searson (1995).

Several definitions deserve mention. An establishment is an economic unit, such as a factory or store, which produces goods or provides services. An establishment is usually a physical location and engaged in one or predominantly one type of economic activity for which a Standard Industrial Classification (SIC) code is applicable. The industry code of an establishment is assigned based on its primary activity, which is determined by the primary product or groups of products produced or distributed (or services rendered) by the establishment.

Employers report employment and wages on an individual establishment basis. Multiple Worksite Reports are used to collect separate employment and wage data for each establishment owned by employers with multiple locations within a state. The Multiple Worksite Reports were instituted as part of the Business Establishment List Improvement Project (BEL breakouts), which was a major initiative conducted jointly by the states and the Bureau of Labor Statistics in 1990 and 1991. The purpose of the BEL breakouts was to get businesses to report their employment and wages at the establishment level rather than the reporting unit level that was state specific prior to the first quarter of 1991. Since the first quarter of 1991 (with the exception of two states that implemented the BEL breakouts in 1992), every multi-establishment employer with ten or more employees in secondary physical locations covered under one UI account has been requested to provide establishment level data.

Employment for a given month is the number of covered workers who earned wages during the pay period which includes the 12th of the month. The employment count includes all corporation officials, executives, other supervisory personnel, clerical workers, wage earners, persons on paid vacations or paid sick leave, pieceworkers, part-time workers, and workers earning wages which are nontaxable under UI because the taxable wage limit has been exceeded. The employment count excludes workers who were on leave without pay or who earned no wages during the applicable pay period because of strikes, work stoppages, or temporary layoffs.

The quarterly UI microdata have information on monthly employment. The LDB publications will use employment in the third month of the quarter as the measure of the establishment's quarterly employment. This policy was selected because comparisons between specific points in time are easier to interpret than are comparisons of quarterly averages. Averaging distorts the timing of when changes in employment occurred, especially changes in employment that occur when an establishment shuts down. Furthermore, monthly employment flows constructed from data reported quarterly might be affected by unknown problems such as quarterly seam effects and other forms of recall bias.

4. Construction of Longitudinal Microdata

There are two pieces of information in the ES-202 microdata that allow for matching establishments across quarters. The first is the SESA ID, which is the UI account number in combination with the establishment's reporting unit (RU) number. The SESA ID is the establishment's unique identifier in the data that the state transmits to BLS. Although the RU number is not used for administration of the unemployment insurance system, the RU number is assigned by the state for BLS purposes of identifying establishments within a multi-establishment employer in that state. The SESA ID is establishment specific rather than location specific, which implies that if an establishment moves across the street in search of a bigger or better location, its SESA ID stays the same.

The second piece of information in the UI microdata used for longitudinal linking is the use of predecessor and successor numbers. The predecessor number is the SESA ID of the establishment that previously owned the establishment in the event of either a change in ownership or a change in reporting

configuration (i.e. a breakout of units). The successor number is the SESA ID of the establishment that will take over the establishment in the event of either a change in ownership or a change in reporting configuration (i.e. a consolidation of units). The term "breakout" refers to a transition from a single establishment employer to a multi-establishment employer, and the term "consolidation" refers to a transition from a multi-establishment employer to a single establishment employer. These breakouts and consolidations may be actual economic events representing business expansions and contractions, or merely administrative reporting changes due to whether or not the business completes the Multiple Worksite Report.

After matching on SESA ID and matching on predecessor and successor numbers both within and across quarters, a third and final step undertaken to link the establishment level microdata across quarters is a probability based match that attempts to identify two establishments with different SESA IDs as continuous. This match is based upon comparing births in the current quarter to deaths in the previous quarter and looking for occurrences such as the same name, the same address, and so forth.

Almost all of the establishments identified as continuous from quarter to quarter are matched by SESA ID. Although the predecessor-successor match and the probability based match links only a small number of establishments, these matches have a significant effect on the number of births and deaths. See Robertson, Huff, Mikkelson, Pivetz, and Winkler (1997) for a more detailed description of the matching algorithm used for the LDB.

5. Researcher Access to the LDB Data

The BLS plans to publish quarterly and annual tabulations of job creation and job destruction. Current plans involve producing these statistics for the entire U.S. economy, by industry, by state, by size, and by age. These tables will be put through a non-disclosure review to insure that we consistently protect the identity of the establishments. These tables, some of which will be published and the others which will be available by request, should satisfy the majority of our customers. However, we do anticipate requests by researchers who want to go beyond our published tabulations and "get their hands dirty" in the microdata.

What is the optimal tradeoff between data confidentiality and data access? The suppliers of the data (the businesses), and BLS as custodians of the data, are concerned about the sensitivity of the employment and wage data being stored on the same microdata record as characteristics such as location, size, and industry that could easily identify the specific establishment. The consumers of the data (the researchers) are eagerly anticipating the construction of the LDB because it will provide a wealth of data never before available which can be used to address important research questions relevant to economic theory, employment and wage policy, and a general understanding of the U.S. economy. As part of the process of constructing the LDB, BLS has been wrestling with the question of how to maximize access to the microdata by legitimate researchers while minimizing the risk of a violation of respondent confidentiality.

Historically, access has only been granted to confidential BLS microdata when authorized by the Commissioner of Labor Statistics for a statistical or research purpose that furthers the mission and function of BLS. Although some confidential data from the National Longitudinal Survey of Youth (NLSY) and the Census of Fatal Occupational Injuries (CFOI) are made available to outside researchers under special agreements, even these data do not contain all the available information. With the exception of the NLSY and the CFOI, most previous access to confidential BLS microdata by outside researchers has been through the ASA/NSF/BLS senior research fellows program. Researchers in the ASA/NSF/BLS program must be affiliated with an organization, and a high official in the organization (a Dean of a university or a Vice President of the organization) signs the agreement with BLS committing the organization to abide by BLS confidentiality policy. The researchers sign BLS non-disclosure affidavits pledging to protect the data and not release the data to anyone.

With regard to the large expected demand for access to the LDB microdata, perhaps the two most important questions are who will be able to use the data, and under what conditions? As of this writing in summer 1997, no specific access policies regarding the LDB have been implemented. Although still in the preliminary stages, BLS is exploring various methods that could be used to establish the Federal Reserve Banks as regional access centers for the LDB microdata. Whatever the details are regarding how researchers access the data, BLS is expected to incur costs. Unlike the Center for Economic Studies at the Census Bureau, BLS does not currently charge user access fees. However, as has been demonstrated in the past both within BLS and in other federal agencies, serious research based on our data has benefits in that it often improves the quality of published statistics.

Two other issues relevant to the debate between confidentiality and access are also worth mentioning.

First, it is assumed that researchers will have access to the actual microdata records, with no masking of data and with no censoring of variables. Second, just as access to the longitudinal database will require a proposal review, the research resulting from the access will need to be reviewed for potential breaches of confidentiality before it is disseminated. The statistical techniques of disclosure avoidance are an active area of research at BLS.

6. Research results

As part of the LDB development, I have been using a sample of ES-202 microdata to analyze net employment growth, job creation, and job destruction. I report the results here, focusing on differences across various sectors of the economy.

The microdata used here are from the state of West Virginia from the first quarter of 1990 to the first quarter of 1995. As mentioned previously, the BEL breakouts were introduced in the first quarter of 1991, with two states being developed in 1992. However, eight states were reporting at the establishment level as of the first quarter of 1990, and using data from any of these eight states adds eight extra quarters to the research data used in this project, thus maximizing the "T" in a panel of size NxT. Of these eight states, several were too small for meaningful analysis, and several were too large for my computing capabilities. West Virginia was the largest of the eight states that would fit within my computing capabilities.

Before linking the data across quarters, I excluded private household (SIC 8811), government, and agricultural establishments. I then matched the 21 consecutive quarters of West Virginia UI microdata (1990:1 to 1995:1) by SESA ID. I have also performed a match by UI account (ignoring the RU number) in order to identify breakouts and consolidations not reported by predecessor or successor numbers. On average, 93.4 percent of establishments in two consecutive quarters matched in this first step. Any establishment that did not match on SESA ID was then matched by predecessor and successor numbers. This second step matched approximately 8.3 percent of those eligible, thus resulting in an average quarterly match rate of 93.9 percent for establishments appearing in two consecutive quarters. Finally, I removed all occurrences of zero and imputed employment at birth and death, and I have defined four consecutive quarters of zero or imputed employment as a death.

The reason for removing all occurrences of zero and imputed employment at birth and death is motivated by the following question: should births and deaths be defined by the first and last filing of the UI tax form, or should births and deaths be defined by the first and last appearance of positive employment? Births and deaths are not identifiable events, but rather are processes of discrete steps such as the gestation period when the idea for a new business is formed, the formal application for UI eligibility, and the actual hiring of employees. Defining births and deaths as the first and last quarter of positive employment is necessary when analyzing job flows, because we want to identify changes in employment corresponding to specific points of the establishment's life cycle, rather than the administrative process of activating or de-activating the UI account which is often independent of any job flows. In the longitudinal West Virginia data, 21.9 percent of new UI accounts have zero or imputed third month employment their first quarter of reporting, and 81.4 percent of deactivated UI accounts have zero or imputed third month employment in their final quarter of reporting. The reason that a zero employment value appears when a new UI account is first reported is that business owners probably apply for UI eligibility before actually hiring their first employee. Similarly, reporting a zero employment value when the UI account disappears undoubtedly occurs when a business owner reduces his employment levels to zero but keeps his UI account active by reporting zero employment because he anticipates starting up the business again when economic conditions improve.

The motivation for defining four consecutive quarters of zero or imputed employment as a death is to help define births for those establishments that have a left censored stream of reported zero employment levels, and to help define deaths for those establishments that have a right censored stream of reported zero employment levels. For example, in the West Virginia microdata, we observe 1123 establishments (1.7 percent of the total sample) with at least four quarters of zero employment before the panel is right censored following the first quarter of 1995. In order to implement this four quarter rule, we must delete the first and the last three quarters of the panel. The final longitudinal West Virginia dataset has approximately 35,000 operating establishments in each quarter from 1990:4 to 1994:2.

The research reported in this section is an integral part of the development of the LDB, but should in no way be considered a sneak-preview of the future published numbers. I have not matched predecessor and successor numbers within, rather than across, quarters, and I have not made use of the probability match that would attempt to further match births and deaths based upon characteristics such as name and address. Because of this, the longitudinal dataset used in the research reported here will not duplicate the data

for West Virginia in the forthcoming LDB. However, the research of Robertson, Huff, Mikkelson, Pivetz, and Winkler (1997) suggests that any discrepancies will be small.

We now turn to empirical estimates of the decomposition of net employment growth into job creation and job destruction. Notationally, let E_t denote aggregate employment in quarter t, let e index establishments, define S^+ as the sector of expanding and opening establishments, and define S^- as the sector of contracting and closing establishments. The average quarterly net employment growth rate is written as

$$\left(\frac{1}{14}\right)\sum_{t=91:1}^{94:2}\left[\frac{E_t - E_{t-1}}{\{E_t + E_{t-1}\}/2}\right],$$

the average quarterly job creation rate is defined as

$$\left(\frac{1}{14}\right)\sum_{t=91:1}^{94:2}\left[\sum_{e\in S^+}\left(\frac{E_t^e - E_{t-1}^e}{\{E_t + E_{t-1}\}/2}\right)\right],$$

and the average quarterly job destruction rate is defined as

$$\left(\frac{1}{14}\right)\sum_{t=91:1}^{94:2}\left[\sum_{e\in S^-}\left(\frac{\left|E_t^e - E_{t-1}^e\right|}{\{E_t + E_{t-1}\}/2}\right)\right].$$

As is evident in the above equations, I follow Davis, Haltiwanger, and Schuh (1996) by using the mean of employment in the current and the previous quarter as the measure of employment in the denominator when converting employment levels into rates.

Measured on a quarterly basis, net employment growth and job creation and job destruction rates are presented in Table 1 for the entire state of West Virginia (excluding private household, government, and agricultural establishments) and for each major industry division. Note that whereas job creation refers to opening and expanding establishments and job destruction refers to closing and contracting establishments, net employment growth also includes the net effects of breakouts and consolidations. Therefore, the job creation and job destruction rates in Table 1 may not precisely add to the net employment growth rate (for example, the net employment effect of breakouts and consolidations when measured on a quarterly basis is 0.1 percent).

For the state of West Virginia as a whole during the early 1990s, employment in the average quarter is growing by 2,341 jobs. Converted into percentage terms, employment is growing on average at the rate of one-half of one percent per quarter (the 0.5% statistic in the upper right corner of table 1). This average net quarterly employment growth can be decomposed into a quarterly job creation rate of 8.4 percent and a quarterly job destruction rate of 8.0 percent. These statistics imply that the average expanding establishment (including births) is growing by 8.4 percent per quarter, and the average contracting establishment (including deaths) is declining by 8.0 percent per quarter.

Looking at the industry statistics in Table 1, quarterly job creation and job destruction rates are lowest in manufacturing, at 4.9 and 5.8 percent respectively. These quarterly job creation and job destruction rates in manufacturing are quite similar to those reported by Davis, Haltiwanger, and Schuh (1996) -- 5.2% and 5.5% respectively, and are also quite similar to those reported by Anderson and Meyer (1995) -- 5.8% and 6.2% respectively. While we should not expect to replicate the job creation and job destruction rates found by these other studies, I find the similarity amazing in light of the different samples and the different time periods used in these various studies.

Table 1: Net Employment Growth Rate
Job Creation and Job Destruction Rates
Quarterly Averages
West Virginia 1990:4 - 1994:2

	%Emp	J.C.	J.D.	N.E.G.
Total				0.5%
		8.4%	8.0%	
Mining	6.2%		10.8%	-1.9%
		8.8%		
Construction	5.9%	23.3%	21.0%	2.2%
Manufacture	16.9%			-0.4%
		4.9%	5.8%	
TPU	7.1%			0.3%
		6.4%	6.0%	
Whole Trade	6.1%			0.0%
		6.4%	6.3%	
Retail Trade	24.1%			0.3%
		8.7%	8.5%	
FIRE	5.0%			0.2%
		6.6%	6.4%	
Services	28.5%			1.4%
		8.2%	6.8%	

%Emp = Percent of total employment in the industry (averaged over 1990:4 to 1994:2). Job Creation (J.C.) - Job Destruction (J.D.) = Net Employment Growth (N.E.G.). Statistics in table may not add precisely because of rounding, because the industry "Not Elsewhere Classified" is

not reported, and because employment effects of breakouts and consolidations are not reported.

The job creation and job destruction statistics in manufacturing are slightly below the quarterly job creation and job destruction rates of 6.0 to 6.6 percent observed in TPU, wholesale trade, and FIRE. Services and retail trade have somewhat higher quarterly job creation and job destruction rates, and mining has the highest quarterly job creation and job destruction rates, at 8.8 and 10.8 percent respectively, of any industry except construction.

We see immediately in Table 1 that the job creation and job destruction rates in construction are outliers. This was also found by Anderson and Meyer (1995). The explanation is the large seasonality of employment in the construction industry. In the first and fourth quarters when construction employment declines, the quarterly job creation rate is roughly 16.6 percent and the quarterly job destruction rate is roughly 27.3 percent. In the second and third quarter when construction employment expands, the quarterly job creation rate is roughly 30.1 percent and the quarterly job destruction rate is roughly 14.7 percent. Many of the other industries exhibit seasonal variation in their quarterly employment levels, but not nearly as large (in percentage terms) as the seasonal variation in quarterly construction employment.

7. Discussion and Extensions

The research presented in this paper has only scratched the surface regarding the usefulness of job creation and job destruction statistics. While I have presented the quarterly data by industry, there are many more dimensions by which it is possible to present the data. For example, we can analyze job creation and job destruction at other frequencies such as annual comparisons. Tables by age will shed light on how establishments grow and decline based upon where they are in their life cycle, and tables by size will clarify the current debate about the (net) magnitude of jobs created by small businesses. The job creation and job destruction statistics presented here are means from the distribution of jobs created at expanding establishments and the distribution of jobs lost at contracting establishments -- summarizing these distributions by statistics other than the mean (such as the variance or by using quartiles) will inform us as to whether job growth and job loss are concentrated at just a few establishments or spread evenly across many establishments. The job creation and job destruction statistics can also be decomposed into the jobs gained by births versus expansions and the jobs lost by deaths versus contractions -- Spletzer (1995) concludes that the amount of job creation and job destruction attributable to births and deaths is non-negligible in the short run and is quite high in the long run.

Davis, Haltiwanger, and Schuh (1996) have shown that job destruction rates in manufacturing exhibit greater cyclical variation than job creation rates. In particular, recessions are characterized by a sharp increase in manufacturing job destruction accompanied by a relatively mild slowdown in job creation. As the LDB becomes a part of the regular production cycle of BLS, we will be able to assess the cyclicality of job creation and job destruction rates in all industries. This information should help economists, policy-makers, and the business community develop a more complete understanding of the business cycle.

Finally, the longitudinal database that the BLS is constructing from the UI establishment microdata might provide a first step towards the ultimate goal of a longitudinal linked employer-employee dataset. Along with the total employment and wages that each business files on its UI tax forms, the business also provides quarterly wages for every employee. These employee level data are known as the "wage records." The tremendous potential uses of a longitudinal linked employer-employee dataset are explained in Lane, Burgess, and Theeuwes (1997). Although the BLS does

not currently maintain any wage records, the experiences gained while constructing the LDB will teach us about development, uses and access relevant to a large confidential database should a longitudinal linked employer-employee dataset ever come to fruition.

Disclaimer

All empirical work in this paper is based on the author's calculations and does not necessarily reflect the official position of the BLS. Any views expressed in this paper are those of the author and do not necessarily reflect the policies of the BLS or the views of other BLS staff members.

References

Anderson, Patricia M. and Bruce D. Meyer. 1994. "The Extent and Consequences of Job Turnover." *Brookings Papers on Economic Activity*, pp. 177-236.

Davis, Steven J., John C. Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. MIT Press.

Farmer, Tracy E. and Michael A. Searson. 1995. "Use of Administrative Records in the Bureau of Labor Statistics' Covered Employment and Wages (ES-202) Program." *Proceedings of the Bureau of the Census 1995 Annual Research Conference*, pp. 198-235.

Lane, Julia, Simon Burgess and Jules Theeuwes. 1997. "The Uses of Longitudinal Matched Worker/Employer Data in Labor Market Analysis." Paper presented at the August 1997 American Statistical Association Meetings.

Lane, Julia, David Stevens, and Simon Burgess. 1996. "Worker and Job Flows." *Economics Letters*, pp. 109-114.

Robertson, Kenneth, Larry Huff, Gordon Mikkelson, Timothy Pivetz, and Alice Winkler. 1997. "Improvements in Record Linkage Processes for the Bureau of Labor Statistics' Business Establishment List." Forthcoming, *Proceedings for the 1997 Record Linkage Workshop and Exposition*.

Spletzer, James R. 1995. "The Contribution of Establishment Births and Deaths to Employment Growth." Paper presented at the January 1996 Econometric Society Meetings.

U.S. Bureau of Labor Statistics, 1997. *BLS Handbook of Methods, Bulletin #2490*.