

August, 2005

A COMPARISON OF THE BUSINESS REGISTERS USED BY THE BUREAU OF LABOR STATISTICS AND
THE BUREAU OF THE CENSUS

AUTHORS: Randy Becker, Joel Elvery, Lucia Foster, C.J. Krizan, Sang Nguyen, and David Talan¹

KEY WORDS: Business register, business list.

1. Introduction

The Bureau of Labor Statistics (BLS) and the Bureau of the Census (BOC) each hold separate list files of the universe of U.S. businesses created from independent sources. These business lists serve a number of critical functions including: (1) creating the sampling frames from which surveys and censuses are drawn, (2) providing the data for tabulations of business activity for BOC's *County Business Patterns* and BLS' *Employment and Wages Annual Averages*, (3) providing benchmarks for survey data, and (4) providing aggregates from surveys and censuses drawn from the lists which are combined by agencies such as the Bureau of Economic Analysis to create national economic indicators. As such, similarities and differences in these lists are of critical interest. BLS and BOC have initiated a project to compare and contrast the two registers and to look for opportunities to improve the lists through data sharing.

As a brief historical background, BLS and BOC embarked on a similar comparison project in 1998. The earlier project focused on comparing the two business lists at the micro level for the total number of establishments, employment, payroll, industry classification, geographic location, and business structures. Although substantial progress was made in illuminating similarities and differences in these variables over the two lists, the project was suspended before any analysis could be done on the underlying reasons for the patterns that were found. The subsequent passage of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) in 2002 rekindled interest in the comparison project. Its passage means that there is a greater potential for data sharing between the two agencies, though outstanding issues must be resolved for data sharing to begin. In 2004, BLS and BOC restarted the comparison project.

¹ The authors' affiliations are as follows: Elvery and Talan, BLS and Becker, Foster, Krizan, and Nguyen, BOC. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on methodological and technical issues are those of the authors and not necessarily those of the U.S. Census Bureau. The views expressed do not reflect the official positions or policies of the Bureau of Labor Statistics or the views of other staff members.

The current comparison project will have two major stages of analysis: analysis of published aggregate data and analysis of confidential micro data. The aggregate analysis will compare the number of establishments, employment, and payroll for the reference year 2001 at the national, sectoral, and state levels. The results from this macro analysis are intended to motivate and help guide the micro analysis in the second stage of the project. In that stage, data from the two business registers will be matched at the micro level using a numerical identifier common to both datasets. The micro analysis will follow two main lines of inquiry. The first line of inquiry will build upon the aggregate comparison and will look more deeply at similarities and differences across the datasets for the reference year 2001. The second line of inquiry will evaluate differences in the ability of the two registers to identify births and deaths in a timely fashion for the reference years 2000 to 2003.

This paper reports on preliminary results from the first stage, the aggregate analysis. We compare 2001 data from BOC's *County Business Patterns* to data from BLS' *Quarterly Census of Employment and Wages*. We show that some of the differences between these sources are due to differences in which industries are included in the tabulations. However, important differences remain. The upcoming micro-data comparison work will enable us to pursue these questions further.

2. The Business Registers

This section briefly describes the functions of the two business registers and their underlying data sources. BLS' business register is a list of active business establishments in the U.S. BLS' business register is compiled as part of its Quarterly Census of Employment and Wages (QCEW) program. These data are aggregated and published in *Employment and Wages Annual Averages*. Aggregated *monthly* employment and *quarterly* wage and establishment data are released quarterly through BLS' website and press releases. These data cover the quarter that ended six months prior. BLS' *quarterly* Business Employment Dynamics data, which measure gross job gains and gross job losses, are derived from the register data and are released seven months after the relevant quarter ends. The micro data also serve as a national sampling frame and benchmark for BLS establishment surveys, including the monthly Current Employment Statistics and Producer Price Index surveys. The BLS business register includes *monthly* data on employment and *quarterly* data on total wages, industry classification, and geographic location. For ease of exposition, we will refer to total wages as "payroll".

This paragraph discusses some features of the

BLS business list that are relevant for this paper. For a more complete discussion, see Searson and Farmer (1997). BLS' principal sources of information on businesses are the State Workforce Agencies (SWA). U.S. businesses report to SWAs in compliance with the State Unemployment Insurance (UI) laws. Businesses report quarterly wages and monthly employment each quarter using their state assigned unemployment insurance identification number. New businesses are required to submit an initial status determination form for unemployment insurance purposes. This report includes their federal Employer Identification Number (EIN). Nearly all new employers are identified through the initial status determination process while a few are identified through the UI claims process and UI field auditor investigations. The BLS business register also incorporates information from the Multiple Worksites Report (MWR). The MWR collects monthly employment and quarterly wage information on establishments associated with multi-establishment firms within a state for about 110,000 firms each quarter. The largest employers file consolidated reports that can cover multiple states. It can be difficult to link together the establishments of enterprises with multiple EINs. Finally, industry and geography codes are verified through the Annual Refiling Survey (ARS) which is sent out to one-third of all businesses on a rotating basis so that all businesses are surveyed once every three years.

BOC's business register is also a list of active U.S. business establishments. The main function of BOC's business register is to serve as the basic source from which economic census and survey sampling frames are drawn. These BOC surveys cover a wide variety of economic topics and are published at monthly, quarterly, and annual frequencies. For example, the BOC business register serves as the sampling frame for the Annual Survey of Manufacturers (ASM). The BOC business register also provides data on business activity for the annual *County Business Patterns*. The BOC business register includes *annual* data on the number of employees, payroll, industry classification, and geographic location. The BOC business register also has information allowing establishments to be linked to their parent firm.

The following description of the BOC business register is a brief synopsis of what can be found in Walker (1997). The Internal Revenue Service (IRS) is the initial source of information on businesses for the BOC business register. The IRS provides information on the existence of these businesses, their location, and their operating status from their Business Master File and from Business Income Tax and Payroll Tax forms. The IRS uses the EIN as the primary identifier. In addition to the IRS data, the BOC receives other

administrative record data, primarily concerning industrial classification codes, from the Social Security Administration (SSA). BOC also receives information on location, industry, and ownership type from BLS. The information from industry is the most extensively used, BOC receives information for about 1 million establishments (most of which are small). The business register includes information from the quinquennial Economic Census and the annual Company Organization Survey (COS). The Economic Census covers virtually the entire U.S. economy and collects detailed economic data from all multi-unit establishment companies and large sample of single-establishment companies on a mandatory basis. Information for those small single-establishment companies not included in the mailout comes from IRS and SSA administrative records. The COS is an annual mandatory survey that covers about 52,000 of the more than 250,000 multi-unit companies. Every fifth year, the COS collection is superseded by the Economic Census.

3. Known Differences in the Registers

Before starting our comparison of the published data, it is important to recognize the known areas of differences. While it is not possible to describe all of the differences here, the following discussion highlights some of the known differences. The four main types of differences are differences in collection, scope, data definitions, and reference period.

One collection difference is that the COS is a mandatory survey collected annually for a sample of firms at the national level. Since the COS is collected at the national level, it is possible to link establishments to their parent firms. The MWR is collected quarterly at the state level and covers multiple establishment firms that have a total of 10 or more workers in their non-primary worksites. The MWR is mandatory in 27 states. While the timing of the Economic Census means that there are certain years in which the maintenance of the BOC business register is more comprehensive, the BOC business register is updated continuously. The updating of the BLS business list, which is done quarterly, is less cyclical.

Concerning scope, BLS includes but BOC excludes the following sectors of the private economy from published data: agricultural production, pensions and other funds, trusts and other accounts, postal workers, and private households. BLS has partial coverage of railroad transportation establishments while BOC excludes the industry. BOC excludes most government establishments, but includes hospitals, retail liquor stores, university publishers, and Federal Reserve Banks. BLS covers government establishments for all levels of government, but these are not included in the private industry data we use in this paper. While

both agencies cover most nonprofits, certain nonprofits have the option to remain out of the UI system and therefore are not covered by the BLS business register. BOC classifies auxiliaries as a separate industry while BLS follows the North American Industry Classification System (NAICS) and classifies such establishments by their primary activity.

The definition of an active establishment varies over datasets. BOC defines active establishments as those with positive payroll at *any time* in the year. In contrast, an establishment must have payroll of \$1,500 in any one quarter or at least one employee for 20 or more weeks to be covered initially for UI purposes and thus in the BLS published data. Some states extend coverage further; exact coverage information is available through the Employment and Training Administration. Once covered, establishments remain in the BLS data until they have zero employment for three consecutive quarters. Therefore, the BLS establishment figures include some establishments that have zero employment and payroll in 2001 while BOC excludes establishments with zero payroll. The reason for this difference is that the BLS data is quarterly while the BOC data is annual. The definition of employment in the two datasets is similar, but the definition of payroll differs. The BOC definition includes taxable pension contributions while the BLS definition does not.

There are also differences in the reference periods that might affect our comparison. The BOC reference period for establishment counts is for existence at any point in time in the year. The annual BLS establishment count is the average number of establishments reported by quarter over the year, though quarterly counts are available. The BOC employment number is the number of employees at the establishment for the week that includes March 12th. BLS reports all monthly employment and the annual average employment. We exclusively use the March 12th employment measures to enhance comparability within this article. BLS employment counts include

4. Comparison at the National Level

Even with these known differences, it is interesting to compare statistics derived from each business list. We compare the number of establishments, employment, and payroll from each agency for 2001. We focus on 2001 because it is the primary reference year for the planned micro-data comparison project. The BLS statistics are from the 2001 *Quarterly Census of Employment and Wages* data available through BLS' website and the BOC statistics are from the 2001 *County Business Patterns* data available through BOC's website. We look at percentage differences between the sets of statistics both as BLS relative to BOC and BOC relative to BLS.

We present two sets of numbers for comparison: one with no adjustment for coverage differences and the other with the best possible adjustment using published data. The industries covered by *County Business Patterns* data are a subset of the ones covered by *Quarterly Census of Employment and Wages*, so we restrict the BLS data to be more comparable to the BOC data. To do this, we subtract the following industries from the private sector BLS data (NAICS in parentheses): Crop and Animal Production (111 and 112); Rail Transportation (482); Postal Service (491); Insurance and Employee Benefits Funds (5251); Trusts, Estates, and Agency Accounts (52592); and Private Households (814). We also add government sector employment in hospitals (622) and retail liquor stores (4453). The remaining differences in coverage are due primarily to what establishments are covered by each agency rather than what industries or ownership categories are covered.

We look first at the data that adjusts for coverage differences only by restricting the BLS data to the private sector. These figures are in columns 1 through 3 of Table 1. The most substantial difference between the two agencies is in the count of establishments. The BOC count of establishments is

Table 1: National-level comparisons of BOC and BLS published data with and without scope adjustments, 2001

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	BOC	BLS	(2 - 1)/1	(1 - 2)/2	BLS	(5 - 1)/1	(1 - 5)/5
Establishments	7,095,302	7,724,967	8.87%	-8.15%	7,213,611	1.67%	-1.64%
Employment	115,061,184	108,932,804	-5.33%	5.63%	108,916,710	-5.34%	5.64%
Payroll (millions)	3,989,086	3,952,152	-0.93%	0.93%	3,972,605	-0.41%	0.41%
Adjusted scope?	--	No	No	No	Yes	Yes	Yes

Source and notes: See text.

workers covered by the UI system, which means that interns and work-study employees may not be counted even in covered establishments. BLS and BOC report payroll as the total annual value.

7.095 million while the BLS count is 7.725 million. This difference of about 630,000 establishments translates to 8-9 percent of the total number of establishments for either register. Although BOC has

fewer establishments than does BLS, it has more employment than does BLS. The BOC measure of employment is 115.1 million and the corresponding BLS measure is 108.9 million. The difference between the agencies' employment figures is 6.2 million which translate to about 5-6 percent of total employment on either register. The BLS payroll figure is about \$37 million lower than the BOC figure, which is less than one percent of either agency's payroll figure.

We next consider the figures adjusted for difference in scope, presented in Columns 1, 5, 6, and 7 of Table 1. Controlling for scope brings the BLS establishment count to 7.213 million. This brings the difference between the two agencies' establishment counts down to about 118,000 establishments. Controlling for scope reduces the BLS employment count by approximately 16,000 and increases the difference in employment counts slightly. The difference between the agencies' employment figures is 5.64 percent of the BLS employment count and 5.34 percent of the BOC employment count. Finally, controlling for scope differences increases BLS payroll to 3.97 trillion and reduces the difference in the agencies' payroll figures to 0.41 percent of payroll.

We find that making some simple adjustments for what industries are covered by each agency brings the statistics from the two sources closer together. This reduces the difference in establishment counts by 80 percent and reduces the difference in payroll by over half. However, the difference in employment rises by 0.2 percent. It is interesting to note that the BLS establishment count is higher than the BOC count, while the BLS figures for employment and payroll are lower than the BOC figures. These results are similar to what Walker (1997) found comparing similar data from 1994. Most likely, this reflects differences in treatment of zero payroll establishments and the greater frequency of the MWR relative to the COS, both of which could elevate BLS establishment counts relative to BOC. As the agencies pursue micro-data comparison, we hope to learn more about this pattern.

5. Comparison at the Sectoral Level

We next compare the data across the two business registers at the sectoral level. We use sectoral data that is adjusted for scope in the way discussed above. One limitation of the 2001 sectoral comparison is that the BOC data is classified on a NAICS 1997 basis while the BLS data is classified on a NAICS 2002 basis. The sectors impacted by the NAICS revision are Construction, Wholesale Trade, Retail Trade, and Information. We compare 2-digit NAICS sectors, which should limit the impact of the difference in codes. The sectors excluded from the analysis are: Agriculture, Unclassified, and BOC's Auxiliary category.

The sectoral level data is presented in the graphs in Figure 1. BLS values are on the horizontal axes while BOC values are on the vertical axes. The graphs have labeled points for each sector and a line which represents perfect agreement. For sectors above (below) the line, BOC data is higher (lower) than BLS data. The legend to Figure 1 provides the definitions of the sector labels.

The sectoral level establishment count plots are in the top panel of figure one. Exactly half of the sectors have higher establishment counts in BLS data than BOC data. Of the five sectors that have more than 5 million employees according to either agency and with establishment counts that differ by more than 10 percent (Construction, Manufacturing Wholesale Trade, and Professional Services, and Other Services), the BOC count is higher than BLS count only for Other Services. The largest differences are in Other Services (where BOC has 186,302 more establishments than BLS) and Wholesale Trade (where BOC has 136,293 fewer establishments than BLS). Note that the NAICS revision impacts two of the sectors with establishment counts that differ by more than 10 percent, Construction and Wholesale Trade.

We next compare the employment figures from BOC and BLS, seen in the middle panel of Figure 1. Of the sectors that have more than 5 million employees according to either agency, only Finance and Insurance, Administrative Services, and Other Services have employment that differs by more than 10 percent. The sectors with the top three largest differences are highest in the BOC data. The three sectors and the size of the difference are: Other Services (1.62 million), Administrative Services (1.36 million), and Management (1.15 million).

Finally, we compare the sector level total payroll figures (bottom panel of Figure 1). Payroll is more similar than establishment counts across the two sources. Of the sectors with greater than 5 million employees, the difference in payroll is greater than 10 percent only for Manufacturing, Administrative Services, and Other Services. The two sectors with the largest difference in payroll across the two data sources are Management and Manufacturing. Management payroll is \$94.8 billion lower in the BLS data than in the BOC data while manufacturing payroll is \$86.4 higher in the BLS data.

Half of the sectors have higher employment, payroll, and establishment counts in one agency's data. Mining, Manufacturing, and Transportation are higher in all three measures in the BLS data. Utilities, Finance and Insurance, Management, Education, Arts, and Other Services are the sectors that are higher in the three measures in the BOC data.

6. Comparison at the State Level

We next compare establishments, employment, and payroll at the state level for the reference year 2001. There are a number of reasons that state level data may differ. If the relevance of scope and definitional differences vary across states, we might expect to see heterogeneity in the differences between the two sources across states. This kind of heterogeneity may also occur if there are some differences in the quality of the data collected by the SWAs or if state level data collection provides more accurate state geographic codes than national level collection. We do indeed find differences across the states in how closely the two datasets match each other. While it is true that the national pattern of discrepancies in establishment number and employment are found in most states (i.e. BLS has more establishments, BOC has more employment), the differences vary across states.

Above we show how important it is to control for scope differences when comparing the BOC and BLS business list data. Due to suppression of some data cells, we can not implement all of the adjustments for coverage differences at the state level. The numbers quoted in this section reflect numbers adjusted for all of the differences not involving suppressed cells. This means that in some states we can implement all of the scope adjustments, while in other states we can not. For this reason, we do not include state level graphs or tables. Also, for this discussion we will treat the District of Columbia as equivalent to a state, giving us 51 areas to compare.

We first look at state variation in the establishment counts. In 33 states, the establishment counts are higher in the BLS data than in the BOC data. The average across states in the percent difference in establishment count is 3.9 percent of BOC count and 3.3 percent of the BLS count. There are nine states where the percentage difference in establishment counts is greater than ten percent of the BOC count, all of which have higher counts in the BLS data than the BOC data. In ascending size of the percent difference, these states are: HI, KY, SC, ID, UT, RI, MT, NH, and DC. Most of these states have small establishment counts relative to other states.

Next we look at differences in employment. In 48 states, the BOC has higher employment than BLS. The average percent difference across states is 4.9 percent of BLS employment and 4.7 percent of BOC employment. All four states where the difference in employment is greater than ten percent of BLS employment have higher employment in the BOC data.

In ascending order of the percent difference, these states are: DE, MA, CT, and NJ.

Finally we consider the differences across states in payroll. While at the national level BOC payroll is

higher than BLS payroll, in 27 states BOC payroll is lower than BLS payroll. Averaging across states, the BLS payroll figure is slightly higher and the percent difference is 0.27 percent of BOC payroll and 0.16 of BLS payroll. Alaska, which has higher BOC payroll than BLS payroll, is the only state where the percent difference in payroll is greater than 10 percent.

7. Conclusions

We have found that in 2001, data derived from the BLS business register has more establishments than that derived from the BOC business register. The BOC data has more employment and payroll than the BLS data. We find that restricting the BLS data to the set of industries and ownership classes covered by the BOC data brings the establishment counts and payroll figures close to one another, but the employment differences are not markedly affected. There is heterogeneity when we turn to the sectoral analysis. For example, there are some sectors in which one agency has higher figures for the number of establishments, employment, and payroll. Looking at the state level data, the differences in number of establishments and employment are reflected in almost every state. There are some interesting differences in the size of these discrepancies by state.

The results of the aggregate comparison will be instrumental in determining the direction of the comparison at the micro level. It is not possible to fully understand the differences between the two registers without analyzing the micro level data. It is expected that the results of the micro analysis will prove to be very useful in guiding the two agencies as they improve their individual business registers.

References

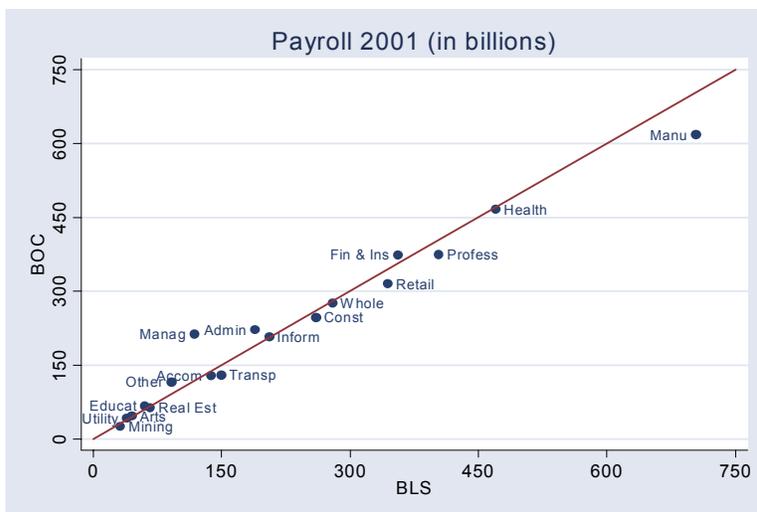
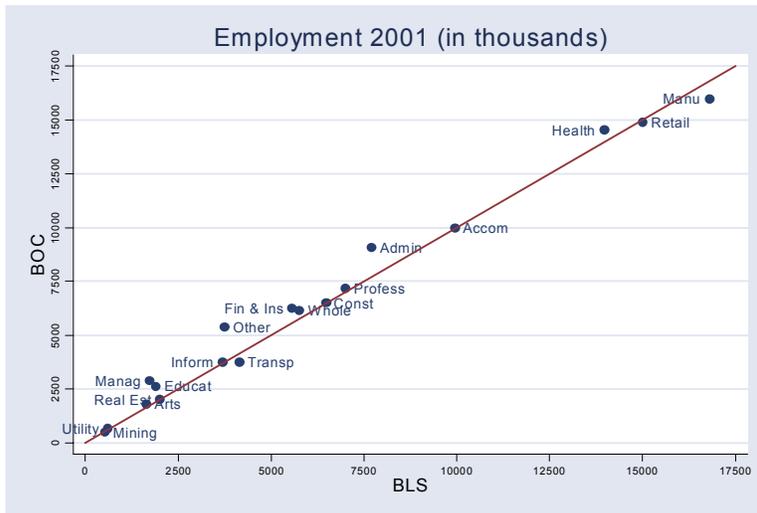
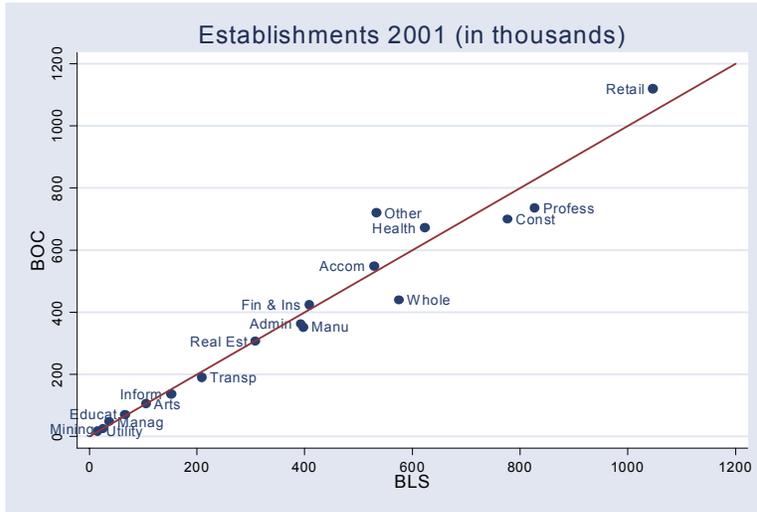
Searson, Michael A. and Tracy E. Farmer. "Quality of the Bureau of Labor Statistics' Business Establishment List as a Sampling Frame." Paper presented at 1997 Joint Statistical Meetings.

Walker, Ed. "The Census Bureau's Business Register: Basic Features and Quality Issues." Paper presented at 1997 Joint Statistical Meetings.

U.S. Bureau of the Census. *County Business Patterns*. Internet distributed files cbp01us.txt and cbp01_ind.txt.

U.S. Bureau of Labor Statistics. *Quarterly Census of Employment and Wages, 2001*. Internet distributed files nt00us01.enb and allsta01.enb.

Figure 1: Comparison at the sectoral level



Legend

<i>Abreviation</i>	Full Name	NAICS Codes
<i>Mining</i>	Mining	21
<i>Utility</i>	Utilities	22
<i>Const</i>	Construction	23
<i>Manu</i>	Manufacturing	31-33
<i>Whole</i>	Wholesale trade	42
<i>Retail</i>	Retail trade	44-45
<i>Transp</i>	Transportation & warehousing	48-49
<i>Inform</i>	Information	51
<i>Fin & Ins</i>	Finance & insurance	52
<i>Real Est</i>	Real estate & rental & leasing	53
<i>Profess</i>	Professional, scientific & technical services	54
<i>Manag</i>	Management of companies & enterprises	55
<i>Admin</i>	Admin, support, waste mgt, remediation services	56
<i>Educat</i>	Educational services	61
<i>Health</i>	Health care and social assistance	62
<i>Arts</i>	Arts, entertainment & recreation	71
<i>Accom</i>	Accommodation & food services	72
<i>Other</i>	Other services (except public administration)	81
Excluded from graphs:		
<i>Agric</i>	Forestry, fishing, hunting, and agriculture support	11
<i>Auxiliaries</i>	Auxiliaries (exc corporate, subsidiary & regional mgt)	95
<i>Misc</i>	Unclassified establishments	99