Business Age and Job Creation: Results from the Business Demography Series

Ron S. Jarmin
Center for Economic Studies
U.S. Census Bureau

Javier Miranda
Center for Economic Studies
U.S. Census Bureau

Census Advisory Committee of Professional Associations Meetings
October 18-19, 2007

ABSTRACT

We describe a new set of annual business statistics generated from the Census Bureau's Longitudinal Business Database (LBD). These statistics provide information on business and employment dynamics including establishment entry and exit and job creation and destruction from 1976 onward and cover all private non-farm businesses in the US. We exploit the LBD's ability to track both establishment and firm activity to generate statistics by firm size and age over a 30-year period. These statistics are unique in their scope and coverage will be a valuable new resource for the research community.

Discussion questions:

1. Size class methodology matters in analyses of net and gross job flows. The different methodologies provide different views of the dynamic process so we are planning on releasing statistics based on several methods. What is the committee's recommendation for the official methodology to use in creating size class statistics? Does the committee think we should have a single official methodology?

2. Due to slight differences in source data and treatment of outliers, there are differences between BDS numbers and published CBP numbers that are based more or less on same underlying data. In particular, there are differences in how one would treat outliers in a longitudinal file as opposed to a single point in time. Does the committee agree with our approach which emphasizes longitudinal consistency over matching other published sources such as the CBP?

We have customized firm size categories based on CBP but there are no published statistics based on Age. We have defined Age categories that made sense based on bin size in order to avoid disclosure problems as much as possible. One consequence of this is that there is more detail for younger firms and less so for older firms. Does the committee think the age categories are appropriate?

Business Age and Job Creation:
Results from the Business Demography Series

Ron Jarmin and Javier Miranda[1]

Prepared for the CACPA
10/18/2007

Preliminary and Incomplete
Do Not Quote of Cite without Permission

[1] Center for Economic Studies, U.S. Census Bureau. Ron.S.Jarmin@census.gov, Javier.Miranda@census.gov.
This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review by the Census Bureau than its official publications. This report is released to inform interested parties and to encourage discussion. Any findings, conclusions or opinions are those of the authors. They do not necessarily reflect those of the Center for Economic Studies or the U.S. Census Bureau.

Measuring Dynamics of US Business using the Census Bureau LBD

INTRO

Business dynamics is a central feature of market economies with broad impacts on labor markets, technical progress and economic growth. The process of Schumpeterian creative destruction is at the heart of the innovative process in modern market economies. Businesses and organizations that develop and/or adopt new and improved products, services and processes grow and displace those that don't. There has been a growing interest in understanding how regulatory and institutional differences across countries affect business dynamics, and in turn, labor markets, innovation and growth. As the competitive environment and business practices change, we also want to understand how firms alter their behavior in response to shocks relative to previous shocks.[2] However, measures of business dynamics are relative newcomers in official economic measurement and have not yet been fully integrated into the broader measurement framework. A recent report from the National Academies (NAS, 2007) reviews currently available statistics on business dynamics, make the case for their importance within the larger economic measurement system and makes suggestions for improvements.

Much of the existing debate about business dynamics surrounds the role of small businesses in the economy. Advocates of small businesses point to statistics showing that small businesses generate the vast majority of jobs in the economy. However, several authors have shown that the relative importance of small vs. large firms in net job creation is sensitive to methodological concerns (the most recent contribution on this topic is from BLS researchers Helfland, Sadeghi and Talan, (2007)). In this paper, we don't tackle this debate directly, but rather argue that confusion over the role of small businesses may stem from not fully accounting for the very important relationship between business age and job creation.

Business Dynamics and the Longitudinal Business Database

In the late 1990s, the Census Bureau's Center for Economic Studies (CES) began development of an economy-wide establishment-level longitudinal database for use in economic research. The development of the Longitudinal Business Database (LBD) was a natural follow-up to its very successful predecessor, the Longitudinal Research Database (LRD). The LRD was used in groundbreaking empirical research on business dynamics by Dunne, Roberts and Samuelson (1989) and by Davis, Haltiwanger, and Schuh (1996), among others. Constructed from respondent-level information in the Annual Survey of Manufactures and the Census of Manufactures, the LRD contained a wealth of information on the activities of manufacturing establishments over time. It was, however, limited to the manufacturing sector.

The development of the LBD was spurred by the need to see if results obtained with the LRD applied to other sectors of the economy, and by the fact that manufacturing's importance as a source of jobs in the economy was decreasing. The source data and basic structure of the LBD are described in Jarmin and Miranda (2002). By utilizing stored "snapshot" files of the Census Bureau Business Register (formerly known as the Standard Statistical Establishment List or SSEL), Jarmin and Miranda were able to construct a longitudinal establishment level dataset

[2] It appears firm's behavior in this last business cycle is quite different from past cycles.

for the private non-farm economy from 1975 to 1999. Subsequent updates have extended coverage through 2005.

The LBD has been utilized in numerous microeconomic analyses including Foster (2003); Jarmin, Klimek and Miranda (2005); and Davis, Haltiwanger, Jarmin, and Miranda (2007). The success of these and other studies has generated substantial interest in public use tabulations from the LBD.

These public use LBD-based tabulations will form the new Business Demography Series (BDS) and are the product of work at the Census Bureau's Center for Economic Studies that has been supported by both the Census Bureau and the Ewing Marion Kauffman Foundation. These will include data on establishment and firm births and deaths, job creation and destruction by firm size, age, and industrial sector, and several other statistics on business dynamics. The LBD-based tabulations being developed have, by design, many elements that are quite similar with the Bureau of Labor Statistics (BLS) Business Employment Dynamics (BED) and the Census Bureau's Statistics on U.S. Business (SUSB) programs. However, the LBD tabulations will offer several important and unique extensions.

1. A long time series. The LBD-based BDS tables cover the period between 1975 and 2005 while the BLS series starts in 1990 and the SUSB series starts in 1989. In this sense the BDS will provide a unique data series that will contribute to our understanding of business demography over the business cycle and longer time horizons. Importantly, analysts will be able to compare firm start-ups, growth, decline, and deaths in periods of rapid innovation and productivity growth to periods with less innovation and productivity growth.
2. Firm and establishment age. These will be the first publicly available tabulations by firm age and by firm size and age. These will enable data users to examine outcomes for cohorts of businesses and see how different cohorts perform over the business cycle. In particular, users will be able examine the role of small and young businesses over a 30-year period.
3. Interactions. The BDS will provide data users the ability to analyze two and three interactions, many which will be unique to the LBD. For example, these will be the first publicly available tabulations of job creation by firm size *and* age.

An Integrated Treatment of Firm Size and Age

For this short paper, we want to focus on one of the more novel features of these new tabulations – their ability to describe business activity across both the size and age dimensions. The source data for Business Demography Series (BDS) is the Longitudinal Business Database (LBD) which is in turn sourced from the Census Bureau Business Register (BR). A key strength of the Census BR is its ability to track corporate structure that comes from both administrative sources and direct data collections. The ability to identify and aggregate all establishments operating in the U.S. under common ownership and to define business size and age accordingly is unique the Census BR.

The LBD links annual snapshots of the Census BR from 1976 to 2005. Included in the LBD is the source information that allows users to track the structure of the firm. Using this firm

identifier LBD records can be aggregated to the firm level for analysis. It's important to remember that the longitudinal linkages in the LBD are at the establishment level. No similar linkage has been developed at the firm level since intertemporal linkages at the firm level can be "many to many" as opposed to the "one-to-one" linkages for individual establishments.

Thus, within any given year it's straightforward to measure the size of firms in the LBD. One simply sums up some establishment level measure of activity, such as employment, across all establishments with the same firm identifier for that year. Measuring changes over-time, or measuring firm age is more difficult.

Measurement of Firm Age

For most firms, the identifier on the LBD does not change from one year to the next. But mergers, acquisitions and other events can lead to changes in firm level identifiers at continuing businesses. This complicates the computation of firm growth and job creation statistics at the firm level. It also complicates the computation of firm age. For firm age statistics in the BDS we follow Becker et. al. (2006) and proxy firm age as the age of the oldest establishment owned by the firm. While this is not a perfect measure, it has two advantages. First, it's easily computed from establishment age information from the LBD. Second, it yields a much more plausible firm age distribution than using the available firm level identifiers to compute the age of the firm. Figure 1 shows the cumulative share of firms, establishments and employment by BDS firm age class in 2005. Over 40% of all firms are age 5 or below. These firms operate about a third of all establishments with employees and employee 13% of all workers in scope for the LBD/BDS. At the other extreme, the left censored category includes those firms in existence at the initial year of the LBD, which account for less than 10% of the firms operating in 2005, own just under 20% of all establishments and employee over 40% of the workers.

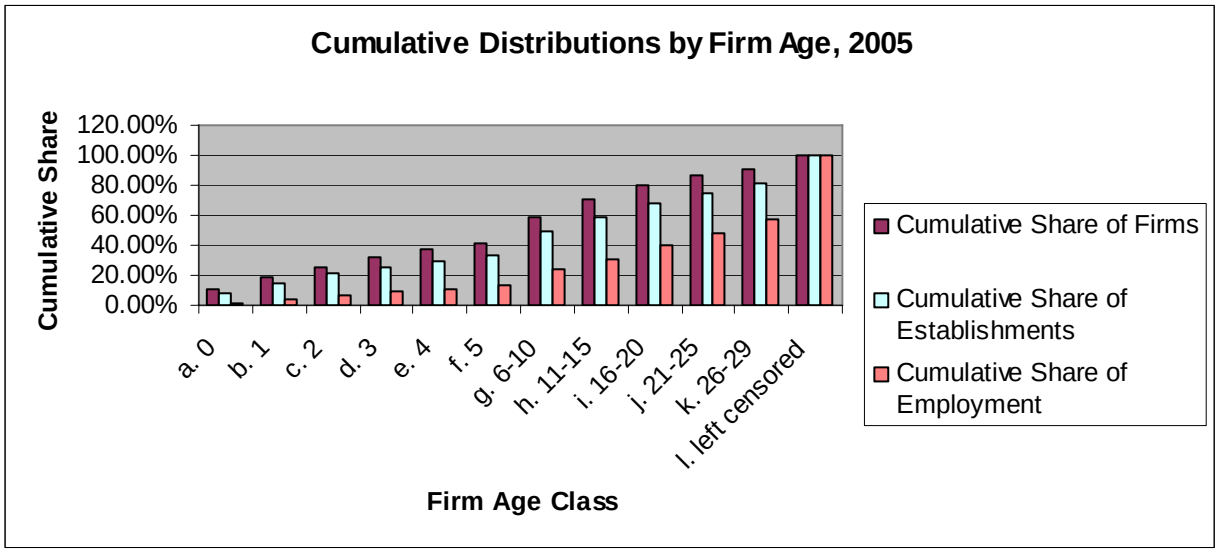


Figure 1

Another important thing to keep in mind is that establishments in the LBD are considered active if they have positive payroll anytime during the year. However, employment is measured as of March 12th. This can generate active establishment records with zero

employment in the LBD. In creating the job creation statistics for the BDS, we use positive March 12th employment to denote active establishments and firms. Our firm age measure is based on the oldest payroll active establishment owned by the firm. In many cases, new single establishment firms begin operations in a particular year after March 12th. That is, the LBD record for that firm will show positive payroll but zero employment. The age measure from the LBD is based on the first year of positive payroll. Thus, our BDS job creation statistics will shift the activity of many births from age zero to age one where positive March 12th employment is first observed.

Measurement of Firm Size

We use the standard size class categories augmented with additional categories to break out the statistics for very large firms. Figure 2 shows the distribution of the number of firms, establishments and employment by firm size, in 2005. In terms of the number of firms, small firms dominate. Firms with fewer than 500 employees – the official Small Business Administration definition – account for more than 99.6% of all firms. However, they own only 84% of the establishments and employ only 52% of the workers in the active LBD universe. In contrast, firms employing more that 10,000 workers account for 0.02% of all firms, own just under 9% of all establishments and employ more that a quarter of the workforce covered by the LBD.

To assign businesses to size classes when computing job creation statistics we use the average size of the business over the two years under consideration. Say, for example, we are describing the net job creation between 2002 and 2003. If a firm grew from 7 to 23 employees between 2002 and 2003 it obviously transitions from the 5-9 to the 20-49 size class. Our methodology would capture the job creation associated with that firm in 10-19 firm size category. We do this to avoid the overstatement of job creation (destruction) at small (large) firms that results when using the initial (terminal) endpoint as the point of reference (see Okolie, 2004). Our method has the added advantage of treating job creation and destruction symmetrically and is easy to implement.

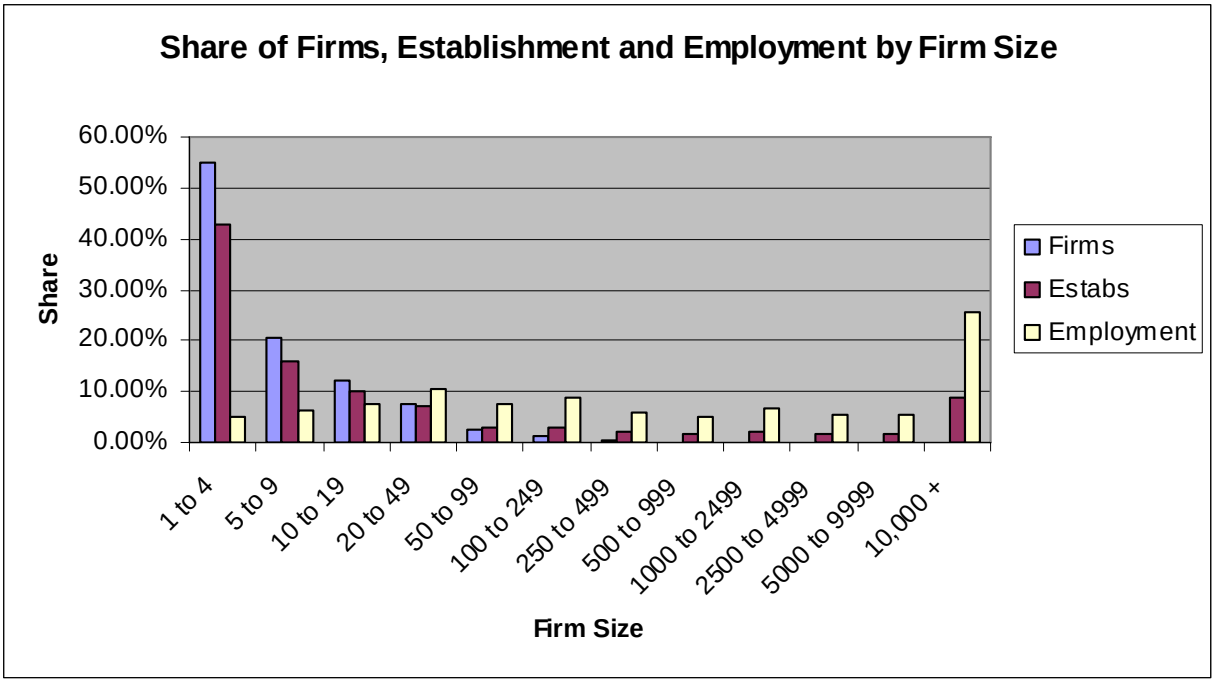


Figure 2

Prototype Job Creation Statistics by Firm Size and Age

We now present just a small sample of the types of statistics to be made available in the BDS. While many data series that release statistics by establishment and firm size break out establishment or firm births and deaths, none look systematically across the age distribution businesses. The BDS will contain statistics on both establishment and firm age. For this paper we focus on firm age since we believe the BDS offers a more nuanced take on the job creation and firm size debate.

Table 1 shows net job creation by firm size and age for the three most recent years available, 2003-2005. Net job creation is equal to job creation minus job destruction for each cell. That is, in each cell, some firms will be expanding and creating jobs and others will be contracting and destroying jobs and still others will have no change at all. The difference between the jobs created at expanding firms and the jobs destroyed at contracting firms yields the net job creation for the cell. The rows and columns sum to the total number of jobs added or subtracted for the private non-farm universe covered by the LBD (very similar to that covered by the Census Bureau's County Business Patterns program).[3]

One of the features that stands out from the table is that job creation is driven largely by new firm entry. This is seen by looking at the row totals in the right hand column. Recall that since our firm age measure is based on the first year with positive payroll, the employment based firm births in the table will be spread across age zero and age one firms. Age zero firms are those born early enough in the year to have positive March 12th employment. Those firms born after March 12th won't show positive employment until age one. So births dominate the age zero and age one categories.

[3] Differences are primarily due to alternative approaches to outlier detection and treatment.

Regardless of the year, firms aged less than two years yield net job creation of approximately 3 million workers. All other age categories yield net job destruction. That is, the number of jobs created at expanding older firms is not enough to counter those destroyed at older firms that are closing or contracting.

Table 1 shows more year-to-year variation in net job creation across the firm size distribution than is typically acknowledged in these kinds of statistics. For instance, in 2003 firms with more that 500 employees shed over 430,000 workers. This was driven by firms with more than 5000 workers. Firms with between 500 and 4999 workers showed positive net job creation.

Examination of the interaction of firm age and size reveals much richer patterns than previously realized. Specifically, looking in the interior of the table shows a complex relationship between firm size and age and net job creation. Small older firms consistently show net job destruction. Larger older firms, however, exhibit both net job creation and destruction with substantial year-to-year variation. Indeed, it appears that, given the consistency of the contribution of very young and very small businesses over each year, what happens at larger, older firms is what drives year-to-year differences in economy wide net job creation,

Table 1. Net Job Creation by Firm Size and Age

Firm Age	a) 1 to 4	b) 5 to 9	c) 10 to 19	d) 20 to 49	e) 50 to 99	f) 100 to 249	g) 250 to 499	h) 500 to 999	i) 1000 to 24	j) 2500 to 49	k) 5000 to 99	l) 10000+	All
2003													
	Firm Size												
0	339,847	144,122	118,152	124,100	78,268	89,849	38,351	30,727	29,004	5,425	0	0	997,845
1	623,418	328,112	281,167	253,927	106,001	71,709	36,816	27,588	24,169	3,283	10,539	-32,836	1,733,892
2	-99,061	-8,281	3,770	12,712	11,021	12,107	13,094	10,717	6,465	9,570	D	D	-27,886
3	-81,869	-22,618	-11,008	4,146	6,326	5,306	2,630	322	13,485	-5,427	D	D	-88,708
4	-64,879	-23,636	-14,366	-5,189	1,371	6,313	5,068	-2,083	-32	18,392	D	D	-79,040
5	-53,795	-20,872	-13,467	-2,356	8,085	12,157	10,741	8,100	6,129	21,033	1,725	-14,759	-37,280
6-10	-177,605	-88,900	-56,824	-21,691	12,002	27,347	27,342	-13,570	-17,256	22,306	5,450	28,239	-253,161
11-15	-89,750	-53,406	-37,850	-19,460	6,541	16,745	6,154	11,300	-6,593	14,953	-11,387	-1,772	-164,524
16-20	-60,659	-41,222	-33,195	-19,797	2,161	15,533	4,529	-10,091	824	27,283	-10,181	41,550	-83,267
21-25	-41,729	-26,320	-25,048	-16,381	1,068	17,734	8,734	-1,982	21,246	11,053	8,310	-9,355	-52,670
26-28	-16,735	-14,592	-19,511	-23,306	-2,594	19,591	16,339	5,139	-15,665	3,954	10,711	-10,318	-46,987
Left Censored	-80,650	-62,418	-58,835	-50,963	-5,237	20,368	-26,119	-30,074	-51,699	-114,536	-129,216	-405,660	-995,036
All	196,536	109,968	132,985	235,743	225,011	314,759	143,677	36,094	10,078	17,289	-100,988	-401,421	**919,730**
2004													
0	365,755	151,612	122,057	130,605	76,169	89,986	19,293	20,546	D	D	0	0	976,023
1	650,399	356,516	304,600	287,442	133,550	91,665	32,276	16,625	12,991	587	3,861	16,306	1,906,817
2	-102,695	-8,155	3,393	8,004	3,511	10,558	11,689	10,249	11,685	D	D	D	-51,762
3	-73,205	-20,989	-8,494	-220	4,159	9,762	-2,559	5,851	-1,133	6,279	-918	20,371	-61,096
4	-60,791	-18,095	-9,353	-740	4,485	6,271	4,224	8,610	D	-3,318	D	D	-68,707
5	-51,906	-19,639	-10,753	-3,730	1,969	-1,139	6,732	9,953	7,085	-1,795	3,886	2,565	-56,773
6-10	-166,726	-78,086	-48,650	-22,987	10,212	9,350	5,975	7,769	8,477	-13,318	1,384	5,871	-280,729
11-15	-90,815	-46,676	-31,367	-22,671	1,570	6,374	6,285	-6,529	17,867	-358	12,058	29,005	-125,258
16-20	-60,317	-34,023	-28,206	-16,512	-6,081	-3,256	5,368	19,978	18,166	4,445	23,821	-32,039	-108,656
21-25	-40,446	-24,775	-21,086	-17,906	-9,856	-2,902	3,354	-6,837	9,965	-13,671	12,957	18,522	-92,681
26-29	-21,614	-17,066	-16,583	-17,510	-8,843	1,202	-6,929	5,725	24,181	10,620	31,926	84,747	69,856
Left Censored	-89,073	-63,873	-58,319	-61,954	-38,384	-10,203	-6,843	252	-25,792	-1,004	-54,951	99,343	-310,800
All	258,568	176,751	197,240	261,822	172,461	207,668	78,865	92,191	86,961	455	36,024	246,895	**1,815,898**
2005													
0	381,751	170,875	140,652	137,625	83,788	69,419	66,129	50,058	59,548	D	D	0	1,159,845
1	660,956	371,916	316,810	294,764	114,035	79,036	46,986	40,095	30,759	D	D	-73	1,955,284
2	-115,391	-17,079	-6,159	-2,217	4,223	7,471	11,357	7,761	806	D	4,558	D	-104,670
3	-93,566	-31,317	-22,082	-14,099	1,293	1,928	12,289	-360	-3,050	D	D	-56	-149,020
4	-68,012	-28,280	-21,875	-8,113	-1,984	7,347	9,251	8,415	352	-3,675	805	1,524	-104,245
5	-60,525	-28,750	-22,471	-10,555	-197	3,667	2,939	4,768	10,003	D	D	D	-101,121
6-10	-209,211	-119,740	-104,043	-81,789	-16,642	7,947	14,887	28,271	-11,398	20,621	-1,789	37,997	-434,888
11-15	-115,695	-80,410	-79,382	-65,676	-20,107	-233	19,944	-884	12,547	17,500	32,843	69,819	-209,735
16-20	-78,737	-62,586	-64,256	-63,993	-17,138	248	19,961	12,451	11,974	2,418	2,064	42,801	-194,793
21-25	-53,561	-46,468	-49,273	-46,996	-18,873	1,743	17,416	5,851	13,075	23,695	-40,528	-25,768	-219,687
26-30	-36,175	-31,801	-40,066	-44,835	-12,471	6,445	12,927	28,550	37,188	31,423	8,963	-23,691	-63,544
Left Censored	-114,063	-84,483	-94,396	-97,039	-47,084	6,942	30,659	38,664	29,492	37,129	51,320	117,275	-125,585
All	97,774	11,877	-46,541	-2,922	68,843	191,960	264,745	223,640	191,295	193,022	84,570	228,901	**1,507,163**

References

Becker, Randy, John Haltiwanger, Shawn Klimek and Daniel Wilson, "Micro and macro data integration: the case of capital," (2006), in Jorgenson, Landefeld and Nordhaus (eds.), A New Architecture for the U.S. National Accounts, University of Chicago Press.

Helfand, Sadeghl and Talen (2007) BLS MLR.

Dunne, Timothy, Mark Roberts and Larry Samuelson, 1989. "Plant Turnover, and Gross Employment Flows in the U.S. Manufacturing Sector," Journal of Labor Economics, 7, no. 1, 48-71.

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

Davis, Steven J., John C. Haltiwanger, Ron S. Jarmin and Javier Miranda, 2007 "Volatility and Dispersion in Business Growth Rates: Publicly Traded vs. Privately Held Firms." NBER Macroeconomics Annual 2006, vol. 21..

Foster, Lucia, 2003, "Establishment and Employment Dynamics in Appalachia: Evidence from the Longitudinal Business Database." CES Working Paper 03-19.

Jarmin, Ron, Klimek, Shawn and Javier Miranda, "The Role of Retail Chains: National, Regional and Industry Results," (forthcoming), in Dunne, Jensen and Roberts (eds.), Producer Dynamics: New Results from Micro Data, available at http://www.nber.org/booksnew/dunn05-1.

Jarmin, Ron S., and Javier Miranda, 2002, "The Longitudinal Business Database", CES Working Paper 02-17.

National Research Council 2007 "Understanding Business Dynamics: An integrated Data System for America's Future" Hatiwanger, John, Lisa M. Lynch, and Christopher Macker, Editors

Cordelia Okolie, "Why Size Class Methodology Matters in Analyses of Net and Gross Job Flows." July 2004 Monthly Labor Review.

****

Appendix: Comparisons to other data sources.

1. County Business Patterns and SUSB

The County Business Patterns program of the US Census bureau is an annual data series that provides economic data by industry and at different levels of geography. This series has been published annually since 1964 and at irregular intervals dating back to 1946. The source for basic data items for the CBP and the LBD are the same; mainly, the Business Register maintained by the U.S. Census Bureau and Census data collections including the COS and the ASM. Given the commonality in data sources and the long time series available for comparison we focus on these data first. Before doing so we discuss some of the characteristics of the County Business Patterns in some detail.

The CBP program has been producing annual statistics on different measures of economic activity since 1964. Its coverage includes most of the private non-farm economy. The CBP's long time series and wide coverage explains its appeal to a wide constituency of users including researchers, policy makers and the business community alike. It is used by the research community to study the economic activity of small areas and to analyze economic change over time; by program areas as a benchmark for statistical series, surveys, and databases between economic censuses. Businesses use the data for analyzing market potential, measuring the effectiveness of sales and advertising programs, setting sales quotas, and developing budgets. Government agencies use the data for administration and planning.

Table 1 presents a detailed listing of the CBP in-scope rules. County Business Patterns covers most of the country's economic activity. The only major exclusions are self-employed individuals, employees of private households, railroad employees, agricultural production employees, and most government employees. This represents the practical totality of the private non-agricultural sector of the economy. We follow the same restrictions in constructing our LBD series.[4]

The comparability of data in the CBP series has been affected over time by definitional changes in establishments, activity status, and industrial classifications (see Census Overview http://www.census.gov/epcd/cbp/view/intro.html). Two changes in particular affect our comparison to the LBD series. The change in the definition of "active" establishments in 1983 and the change in industrial classification systems from the Standard Industrial Classification (SIC) to the North American Industry Classification System (NAICS) in 1998. We are able to bypass these definitional changes in the LBD series although the changes in industrial classification still present unique challenges in the continuity of our series.

Figure 1 compares total employment in the CBP and LBD series. It is immediately obvious upon inspection that the 2 series track each other closely but do not line up perfectly. Both

[4] Code to generate CBP in-scope restrictions is available on request.

the CBP and the LBD use the same underlying data; however, each series imposes its own set of restrictions and edits on the universe of active establishments. Specific differences are difficult to track and get lost in the history of the algorithms and analysts that generate the CBP patterns. The LBD series also imposes its own set of restrictions and edits that are unique to the LBD program including the use of name and address matching to eliminate duplicates and the use of algorithms that identify and clean up incorrect data entries in the employment series. Having said this the differences are not large and range within a band of 1.9 million workers with the LBD typically falling bellow the CBP numbers particularly in the second half of the series.

More relevant perhaps is the comparison of the change in net employment series. Figure 2 presents results from comparing the CBP versus LBD for the whole in-scope economy and by year. Comparison is presented in rates. As before the two series track each other reasonably well. The biggest difference, a 2-point difference, is found at the beginning of the series. The discrepancy can be tracked mostly to differences in education and mining across the two series and are probably due to differences in the scope across series. We can plot the LBD series against the real GDP growth series to isolate growth trends around those years. In general the GDP growth series appears to precede net employment change but trend in the same direction during this time period. With this in mind we still caution users to the discrepancies between the LBD and the CBP for this particular time frame and industry series.

Figure 4 shows the number of active establishments in the LBD and CBP series respectively. The break in the CBP series is due to a change in the way it defined active establishments going from one based on positive employment to one based on positive payroll. The discrepancy in the numbers between the two measures is due to the fact that payroll is measured on a flow bases; it is the sum of payroll for the year, whereas employment is measured at a point in time; it is employment at March 12 of that year. Establishments that start operations after March 12 will show zero employment but positive payroll. We mimic the change in definitions in the LBD to track the CBP series. The chart shows the LBD series consistently below the CBP series in terms of the number of employment active establishments. Differences might be due to edits on the CBP side on the employment variable. In terms of payroll we see… (I have to add payroll to this chart).

The statistics described above show the LBD series and the CBP series track the levels of employment and payroll fairly well. These statistics are readily computed from cross sectional data. However, we are primarily interested in generating business statistics using the longitudinal information in these files. How many jobs are created of destroyed? How many of these jobs come from establishments opening and closing? We use the Bureau of the Census SUSB data series to examine and contrast basic longitudinal characteristics of the LBD data series.

The Bureau of the Census publishes their Statistics of US Businesses (SUSB) series. Data files used in the Statistics of U.S. Businesses (SUSB) program were created from the annual County Business Patterns (CBP) files. The SUSB data files are developed in cooperation with, and partially funded by, the Office of Advocacy of the U.S. Small Business Administration (SBA). Some of the key economic data items they tabulate include number of companies, establishments and their employment. The SUSB program also publishes a

number of measures of business dynamics based on a longitudinal file, the Business Information Tracking Series (BITS) that links establishments from year to year. The file covers the period between 1995 and 2004. Using this file, they are able to tabulate data for establishment births and deaths and employment expansions and contractions.[5]

Figure 5 compares the number of establishment births and deaths in the LBD and SUSB series. Both series track each other's births and deaths well; however, the LBD shows lower numbers particularly for births during census years (those ending in 2 and 7). The smoother LBD is the result of a series of algorithms designed to identify and retime incorrectly timed births and deaths that come about from census processing activities (see Miranda, 2006). The reassignment of births and deaths result in a smoothing of these spikes for those years. Note that the retiming algorithms have a clear impact on the birth series but have a weaker impact on the deaths series. This stems from an asymmetry in the number of incorrect births and deaths (the census does a better job at identifying deaths than births) as well as the relative difficulty in reassigning deaths for multiunit firms. (need to replace these 2 graphs with updated versions).

2. QCEW and BED

The previous section compares the LBD to Census data series that use the same underlying data. In this section we contrast against a series that makes use of a different underlying business register list, the QCEW and the BED. We document a number of discrepancies between the LBD and the BLS series in terms of the number of workers, the number of establishments and the volatility of employment dynamics. Reconciling these discrepancies is beyond the scope of this paper but note that there is a joint BLS-Census project tasked with identifying the source of the discrepancies. Beyond the discrepancies we find similar trends in all the series.

The Business Employment Dynamics (BED) is a set of statistics generated from the Quarterly Census of Employment and Wages, QCEW or ES-202, program. This program is a quarterly census of all establishments under State unemployment insurance programs, representing about 98 percent of employment on nonfarm payrolls. The administrative records are linked across quarters to provide a longitudinal history for each establishment. This linking process allows the tracking of gross job gains and gross job losses from 1992 forward.[6]

Figure 7 compares employment in the LBD and QCEW series. Employment is clearly higher in the LBD series. The difference remains stable at about 3 million workers until 2002 but shots up to about 5 million after that. Differences in levels are to be expected due to significant differences in measured concepts. The bureau of the Census measures March 12 employment while BLS measures only UI covered workers who worked or received pay for the pay period that included the 12th day of the month. The BLS' annual measure is an average of the 12 (one for each month) point-in-time values. Having said this it is unclear at this point what is causing the divergence between the two series.

[5] These files can be downloaded at http://www.census.gov/csd/susb/susb.htm
[6] See http://www.bls.gov/bdm/bdmfaq.htm#1 for information about the BED program.

We also compare the number of establishments in the LBD and QCEW series. Figure 8 shows that here too there are significant differences between the two series. However, in this case differences have been increasing since at least 1990. The source of this growing discrepancy is unclear at this point but there are several possible explanations. Differences may be due to the way establishment units are broken down. For instance, the QCEW will on occasion break up a single physical location encompassing two or more distinct and significant activities. In such cases each activity is reported as a separate establishment, if separate records are kept, and the various activities are classified under different NAICS industries. An increase in the number of such break ups could explain observed trends. Another possibility is the BLS' use of Multiple Worksite Report; which might allow them to better identify the number of establishments associated with small multi unit firms.[7] Changes in the implementation of the MWR across states could then explain a growing discrepancy between the series.

The QCEW and LBD show significant differences in number of workers and establishments they report. These differences are increasing. The question we ask next is whether employment changes track each other well. In other words, do these differences have a bearing on employment flow statistics? Figure 9 shows the results from this comparison between the LBD, QCEW and CBP. For each series we compute the year-to-year net employment change. We find these series track each other well. The divergence in employment levels between the QCEW and the LBD can be observed starting in 1999 with the LBD series generally moving above the QCEW series.

We also want to compare gross flows. Given the differences in the number of establishments and workers differences in the size distribution of establishments as well as gross employment flows are possible. For this we use the longitudinal LBD and BED series.

Before we can compare the two series we have to make some adjustments. The BED and LBD statistics are computed on different frequencies. The BED is a quarterly series whereas the LBD is an annual series. As a result the two series are not directly comparable. Quarterly flows are by necessity much lower than annual flows since changes are captured over shorter periods of time. The sum of four different quarter flow measures will however be much higher since it captures more of the short-term variation in employment. In order to make the two series comparable we use quarterly weights to convert the quarterly flows in the BED series into annual flows.[8]

Figure 10 compares LBD flows against the annualized BED series. We are interested in comparing the changes not the levels so average creation and destruction in the BED series is made to match average rates from the LBD. The annualized BED series is clearly smoother than the LBD series. Both series show a downward trend in the job creation and destruction rates; however, the BED series seems to drop faster after 2001. The source of the discrepancy between the two series is unclear at this point.

4. INDUSTRY, SIZE and AGE.

[7] Most multi-location employers with a total of 10 or more employees combined in their secondary locations are required or requested to complete the MWR.
[8] Weights are estimated by running a regression of quarterly flows on the annual measure. We do this separately using the LRD and also the BED and annualized BED series.

The previous sections compared business dynamic statistics from the LBD to existing series in the public domain including the BLS' QCEW and BED and the Census Bureau's SUSB. This section focuses exclusively on Census data to discuss some of the key variables in the LBD series including industry, size and age.

Users of statistical data want access to data series that are consistent in its classification. Breaks in the series compromise the continuity of the analysis and thus limit their use. Maintaining continuity of the industry codes is particularly daunting for a data series such as the LBD that span a 30-year period. Between 1976 and 2005 the statistical system has introduced multiple industry classifications and revisions including the 1987 SIC revision, the conversion to NAICS in 1997 and the NAICS revision in 2002. Many of these revisions are minor in scope and can be bypassed with a minimum of effort. Others are comprehensive in scope, which make the conversion to a single classification difficult at best. The problem gets progressively more difficult with the level of industry detail we wish to retain since the mappings are often one to many.

For the initial release of the LBD tabs we limit the scope of the challenge by producing a set of statistics at the level of the SIC industrial sector. Even at this highly aggregated level there are a few mappings that cross sectors. We resolve this problem by using longitudinal information to assign SIC codes when ever possible in effect limiting the problem to establishment births. We then use Census bridge SIC codes to resolve establishment births between 1998 and 2001 and assign our own between 2002 and 2005. Census SIC bridge codes are created using a combination of NAICS and SIC sources (see Konschnic et al, 2000). When the mapping is not unique or there are incomplete codes Census employs a random assignment methodology to assign a code. Census estimates 11% of active establishments in the 1999 Business Register have codes that are the result of probabilistic assignment, accounting for 2.5% of 1999 payroll.

With the completion of the SIC to NAICS transition in 2002 Census stopped supporting the SIC bridge codes. Establishments are no longer classified on both an SIC and NAICS codes bases. So our goal is to convert business register NAICS codes to a unique SIC sector code for establishments born after 2001.
We follow a simple methodology to assign SIC sector codes to establishments born between 2002 and 2005. First we identify the best NAICS code available for the establishment. We use Census year codes when complete and available; otherwise we use the modal NAICS code. We then resolve 1-to-1 mappings using a 6-digit NAICS to sector SIC cross walk.

The second step involves reclassifying NAICS auxiliary establishments. This presents particular problems since there is no mapping into SIC codes. In the SIC system, auxiliary establishments (i.e., those establishments primarily serving other establishments of the same enterprise) were classified in the industry of the establishments served. In NAICS, auxiliary establishments are classified according to the services performed rather than the industry served. Specifically corporate, subsidiary and regional managing offices are included in NAICS Sector 55, Management of Companies and Enterprises. In NAICS-based tables from the 1997 Economic Census, all other auxiliary establishments are included in the separate category titled "Auxiliaries, except management of companies and enterprises," and further classified into several broad NAICS industry categories based on the type of service

performed. We get around this problem by using a set of auxiliary codes available in the Business Register that are specifically designed to facilitate this conversion. The mapping takes place at a very crude sector level but is sufficient for our purposes.[9]

Finally, we resolve remaining cases by assigning to the SIC sector with the largest share of employment. We estimate that in 2005 approximately X% of records and X% of payroll are assigned an SIC sector this way.

Figure 11 shows net job creation rates by SIC sector between 1976 and 2005. Results are compared against CBP net job creation rates. Note that the LBD flows are computed from gross job flow measures whereas the CBP numbers are computed from aggregate measures and as a result the two are not perfectly comparable. Despite this fact the pictures show net job creation from the LBD match published CBP numbers very well at the sector level. We find a few exceptions in and around the early and later parts of the 1980's primarily in TCU and FIRE. Turbulence in these sectors from establishment entry and exit might be partially responsible for this.

We cannot compare the LBD SIC series to the CBP after 1997 since publications switched to the NAICS classification. However, to assert that the employment levels are correctly benchmarked figure 12 compares LBD employment levels on a NAICS bases. We find that in general the series track each other with exceptions after 2002 due to the change in the published CBP series from the 1998 to the 2002 NAICS classification.

The use of business age information is critical if we are to measure the activity of young producers as well as their role in dynamic economic processes. The LBD gives researchers the best capability to date to accurately determine the age of establishments in Economic Censuses and surveys. This section discusses its construction and some of its characteristics.

Census survey forms do not routinely capture establishment or firm age information. Previous attempts at collecting this type of information from survey respondents in the late 1970's indicated the information was of limited value. Many establishments failed to provide this information and when provided rounding and recall errors proved to be a concern. The questions were eliminated from the survey forms.

Given the lack of reliable survey response data we take a practical approach to construct establishment and firm age from observed data. We use the LBD's longitudinal links and long time series to generate both establishment and firm age information for all establishments and firms (see Jarmin, 2002). We start by determining the year an establishment first becomes operational in the LBD, let's call it "birth year". Operational status is defined by their reporting positive employment and/or payroll. We then compute establishment age by taking the difference between current year of operation and "birth year". Given that the LBD series starts in 1975 observed age is by construction left censored at 1975. Figure 13 shows the age distribution for establishments and employment for establishments and employment covered by the LBD. Establishments are classified according to their age in the following categories, 0, 1, 2, 3, 4, 5, 6-10, 11-15, 16-20, 21-25, 26-27 and born before 1976. Figure 13 shows that we compute observed age for approximately 74 percent of LBD establishments and 61 percent of LBD employment.

[9] Note this will limit the level of detail to which we can assign these auxiliaries in the future.

The LBD is an establishment level longitudinal database. As such we can track establishment activity and their observed age regardless of ownership or organizational change. We simply track the establishment ID, the LBDNUM. The LBD also contains firm identifiers that allow us to track firm ownership of those establishments. However, unlike the establishment id firm identifiers in the LBD are not longitudinally linked. In other words, ownership changes, reorganizations, mergers and acquisitions, break-ups, divestitures and simple single to multiunit expansions can lead to breaks in the firm identifier. As a result the firm age distribution that results from using the existing firm id leads to unusual shifts in the firm age distribution and spikes linked to Census processing activities.

To get around these issues we follow Becker et al. (2006) to construct an observed firm age. They construct firm age as a dynamic variable that equals the oldest establishment in the firm at a given point in time. Firm age constructed this way displays more plausible age distributions. Figure 14 shows the firm age distribution for establishments and employment covered by the LBD. As before establishments are classified according to age of the firm that owns them into the following categories, 0, 1, 2, 3, 4, 5, 6-10, 11-15, 16-20, 21-25, 26-27 and born before 1976. Firm age constructed this way gives us observed age for approximately 62 percent of LBD establishments and 41 percent of LBD employment. Note that all firms that own an establishment born on or before 1975 have the same age. These older firms tend to be large and, therefore, account for large portion of overall economic activity.

The statistical system has a mandate to produce accurate cross sectional statistics of the state of the US economy. One consequence of this strategy is that our ability to track and measure the activity of small economic units is limited. The LBD gives researchers the capability to track the dynamic activity of establishments of all sizes. This section discusses our use of establishment size in the construction of the BDS.

Sources for the LBD's employment measure include Census collections and administrative records. In both instances establishments are required to provide the number of employees for pay period including March 12. This includes number of employees, both full- and part-time, whose payroll was reported on the organization's Internal Revenue Service Form(s), Employer's Quarterly Federal Tax Return. It includes as employees any persons on paid sick leave, paid holidays, and paid vacations as well as salaried officers and executives of incorporated firms.

The use of a point in time measure of employment as a measure of establishment size has several implications for the way we measure business and employment dynamics. In terms of entry and exit the BDS defines activity based on the existence of positive March 12 employment. Establishments that start operations after this time are considered an entry in the following year. Establishments that cease operations and fail to report employment during the March 12 period are considered an exit that year.

When it comes to computing flows by size class we can use alternative definitions of establishment and firm size based on our choice of classification year. Each yields a different interpretation of the employment dynamics process. For instance, we can use initial year to classify an establishment into a size class. In this case as establishments switch size categories job creation and destruction statistics will emphasize smaller units as the source of that growth. We can also use end year to classify an establishment. In this case job creation

and destruction statistics will emphasize larger units as the source of the growth. Both these measures suffer from what is known as "regression fallacy" or "regression to the mean" bias. The fallacy arises because temporary increases/decreases in employment tend to reverse themselves so that firms that are large in the base year tend to contract while firms that are small tend to increase. The fallacy also holds when using end year.

A related problem in choosing initial or end year to calculate rates is that the resulting statistics are not symmetric. An example helps illustrate this problem. Suppose employment increases from 1 to 2 and then declines back to 1. A growth rate that uses initial year employment in the denominator would result in a 100-percent increase followed by a 50-percent decrease.[10] Asymmetry is also a characteristic of using end-year employment to compute rates.

An often use alternative is to use the average of the two periods in the computation of growth rates. This measure has become standard in the literature because it avoids the regression fallacy when examining size class statistics. This measure provides other desirable characteristics. First, this measure is symmetric with respect to employment increases and decreases. For instance, if average employment were used in the denominator, the growth rate in the above example would be a 67- percent increase [(2 – 1)/1.5 = 0.67] followed by a 67-percent decrease. The use of the average in the denominator leads to rates that are equal in magnitude, but opposite in sign. Second, the rates are bounded between –2 and 2 affording an integral treatment of births, deaths and continuers. Third, this growth rate measure is identical to the conventional measure (change divided by initial employment) up to a first-order Taylor Series approximation and identical to a log change measure up to a second-order Taylor Series approximation.[11]

One possible drawback from using average employment when computing growth statistics is that this methodology allocates the jobs gained during expansions and the jobs lost during contractions to only one size category. For instance, if a firm grows from 2 to 8 employees the establishment will be classified in the 5 to 9 category and all growth of 6 employees will be allocated to the 5 to 9 category. The drawback from doing so is that we fail to credit or allocate any employment change to the initial class category, 1 to 4.

Given there are multiple ways to compute class-size statistics the Census Bureau will release statistics based on initial-year sizing as well as mean sizing.

10 See Davis et al, 1999 for a description of these issues.
[11] See Davis and Haltiwanger, 2001. Journal of Monetary Economics 48 (2001) 465–512

Table 1 CBP In-Scope Restrictions

•Sectors:	–agricultural services, forestry, and fishing;
	–mining;
	–construction;
	–manufacturing;
	–transportation and public utilities;
	–wholesale trade;
	–retail trade;
	–finance, insurance, and real estate;
	–services.
•Exclusions:	–self-employed people,
	–domestic service workers,
	–railroad employees,
	–agricultural production workers,
	–most government employees,
	–and employees on ocean-borne vessels or in foreign countries.
•Establishments:	–The CBP series excludes governmental establishments except for liquor stores (SIC 592), wholesale liquor establishments (SIC 518), depository institutions (SIC 60), federal and federally sponsored credit agencies (SIC 611), and hospitals (SIC 806).
•Employment:	–full- and part-time March 12 employees.
	–Included are employees on paid sick leave, holidays, and vacations;
	–not included are proprietors and partners of unincorporated businesses.
•Payroll:	–Total payroll includes all forms of compensation, such as salaries, wages, reported tips, commissions, bonuses, vacation allowances, sick-leave pay, employee contributions to qualified pension plans, and the value of taxable fringe benefits.

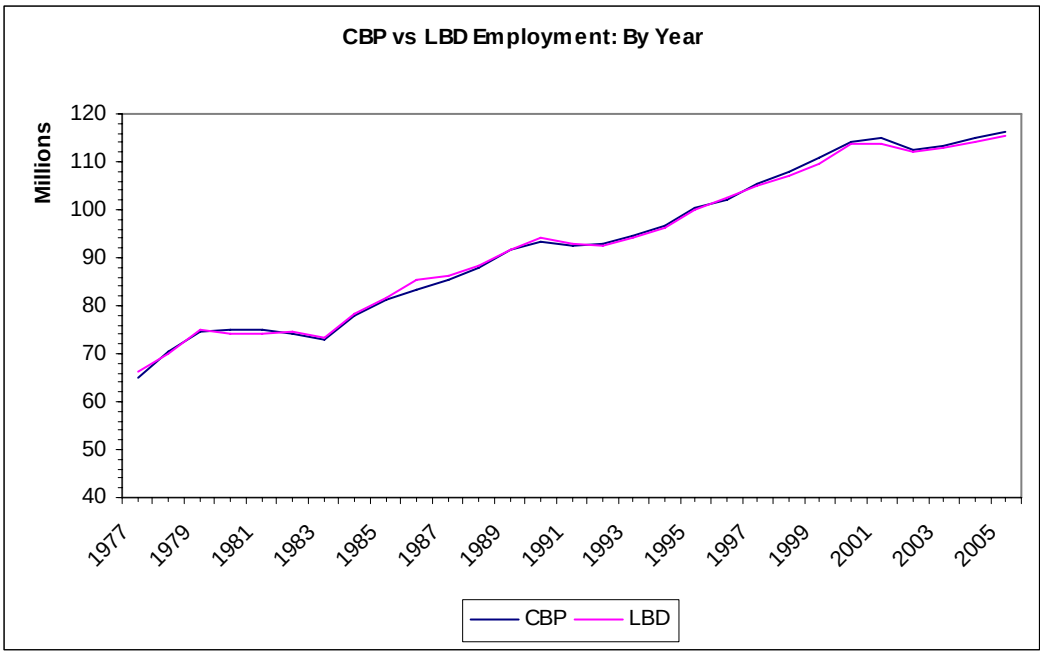


Figure 1

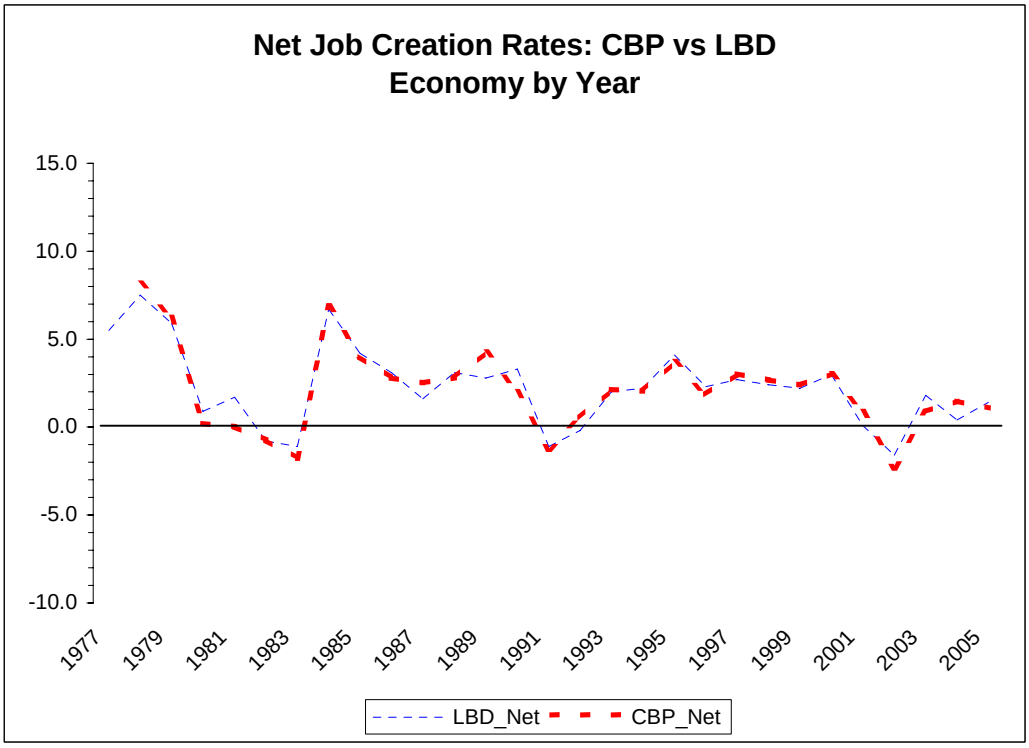


Figure 2

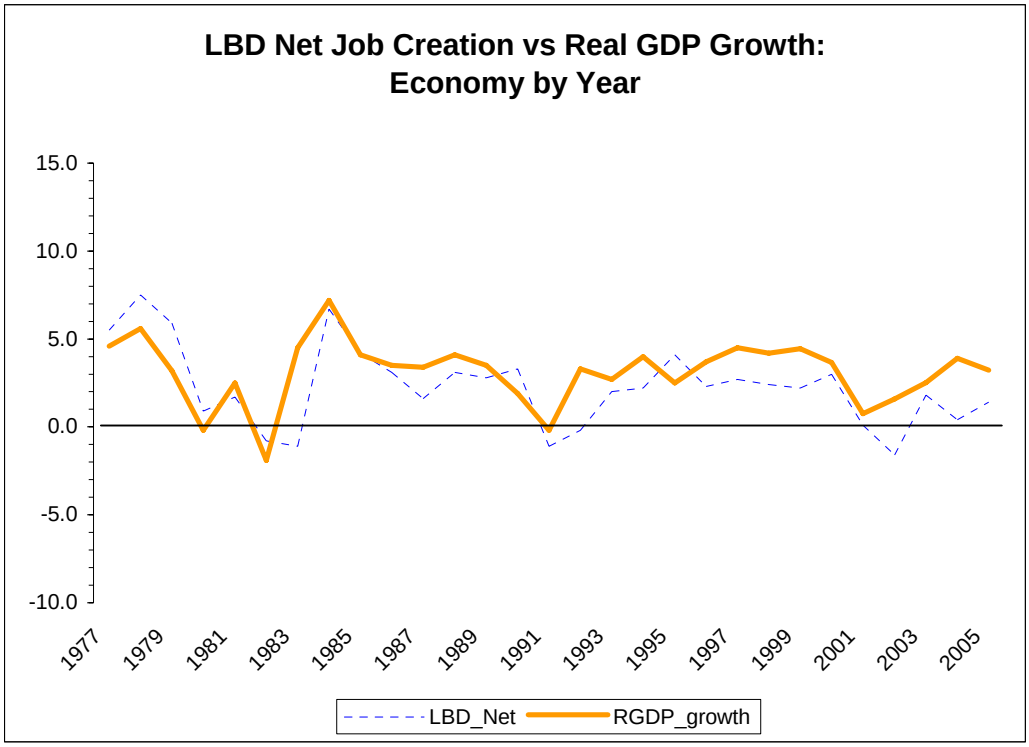


Figure 3

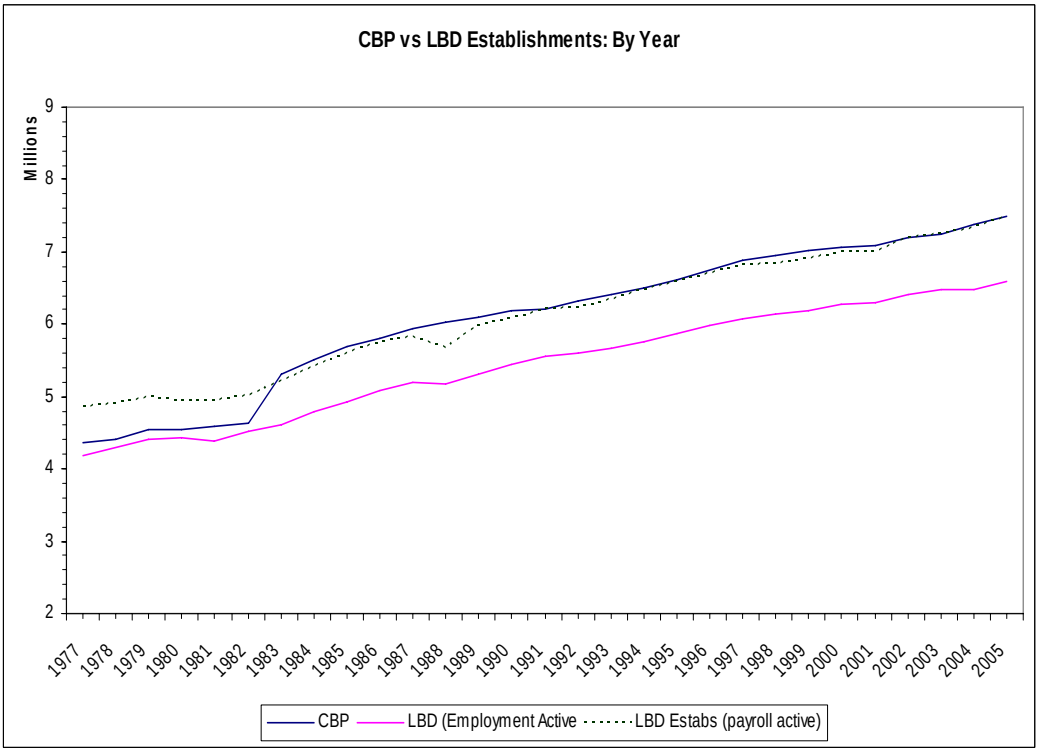


Figure 4 I'm redoing this with new data

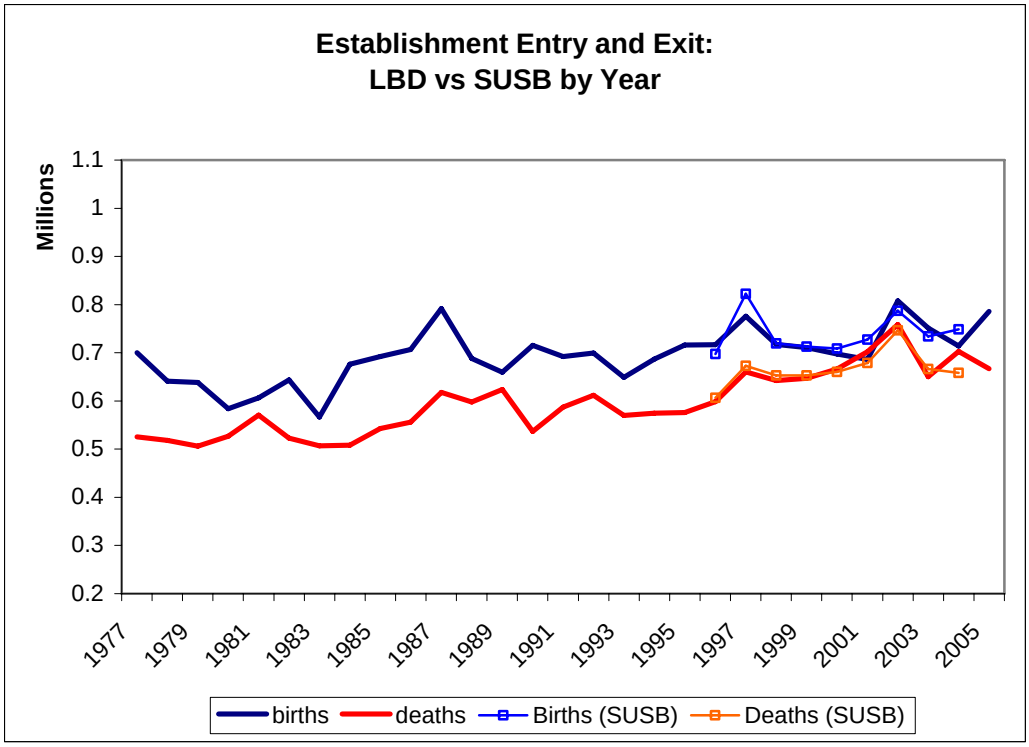


Figure 5

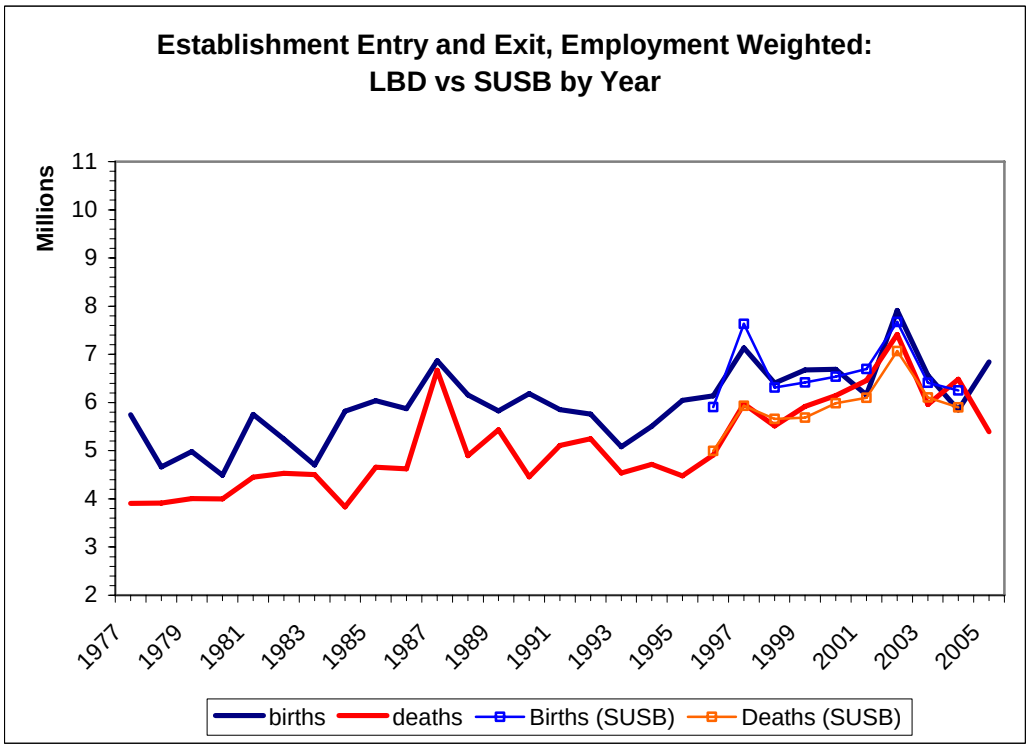


Figure 6

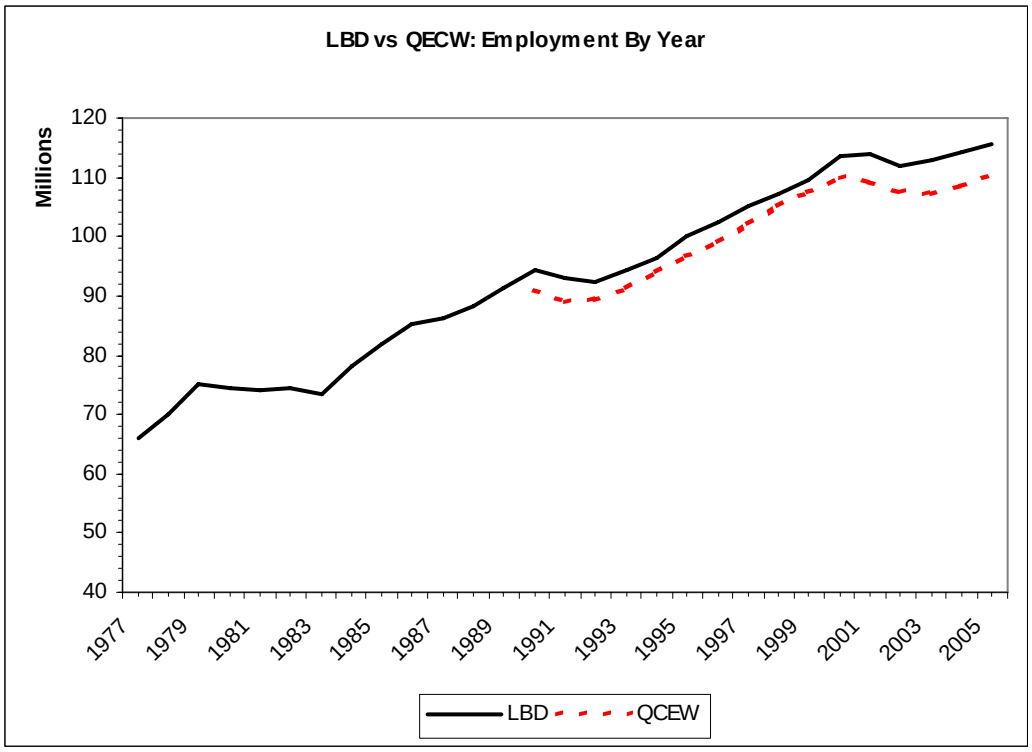


Figure 7

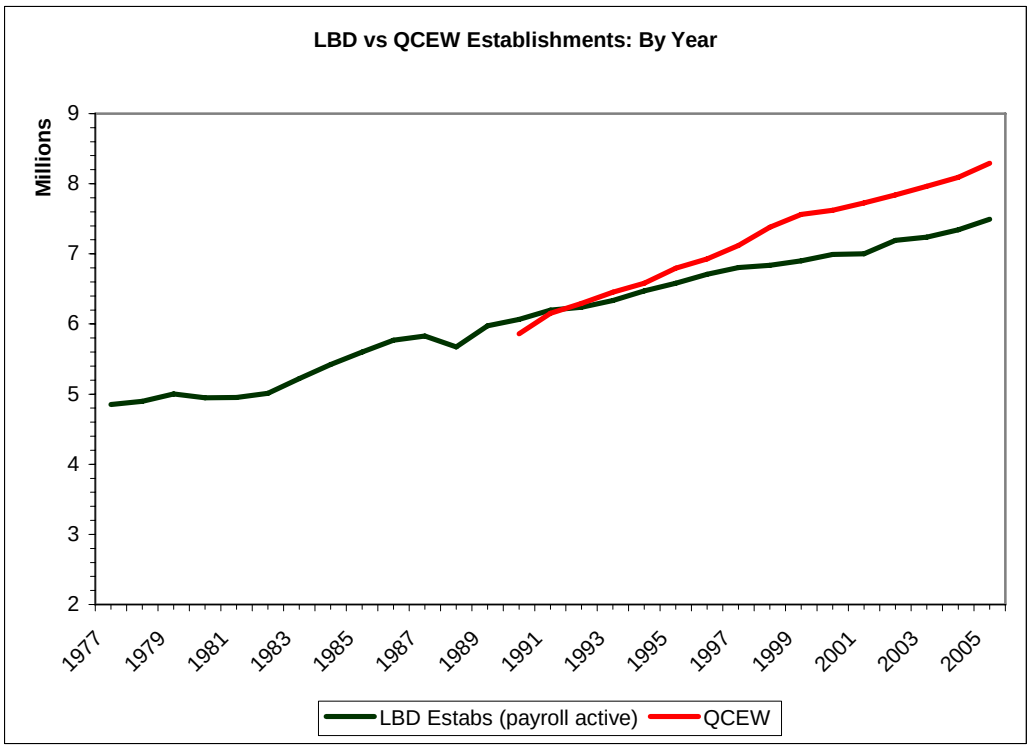


Figure 8

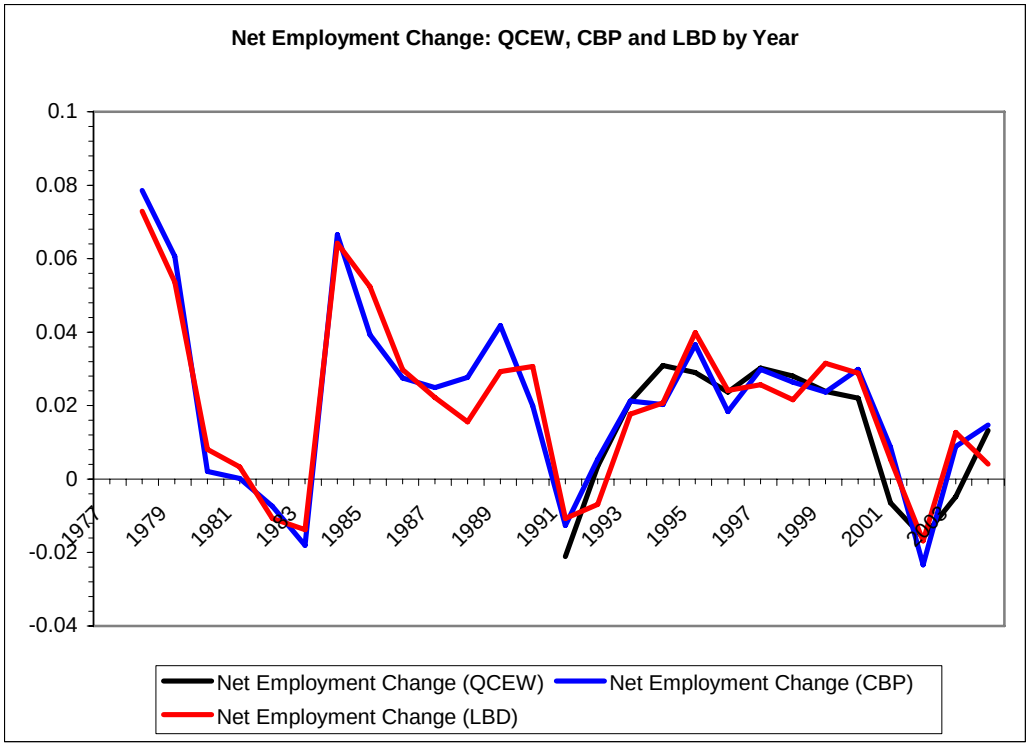


Figure 9

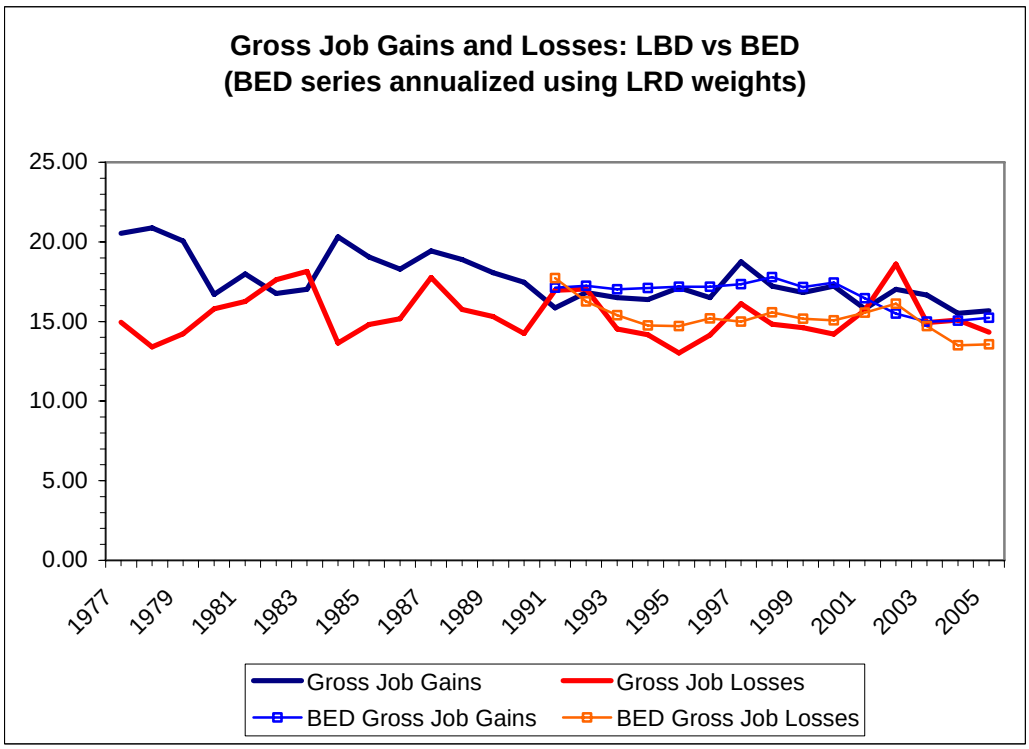
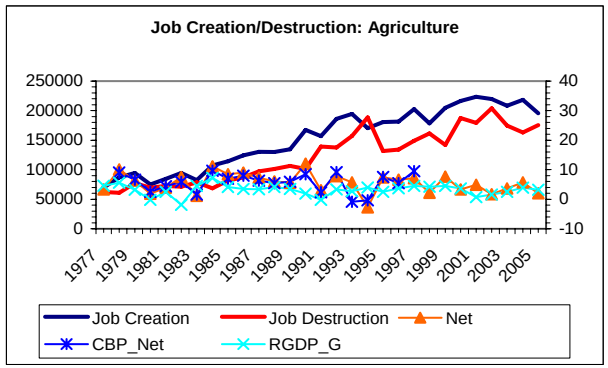
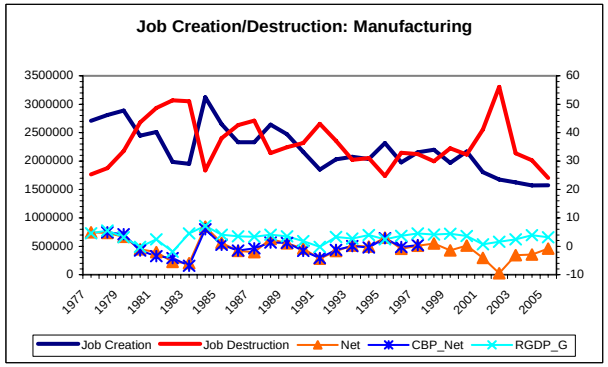
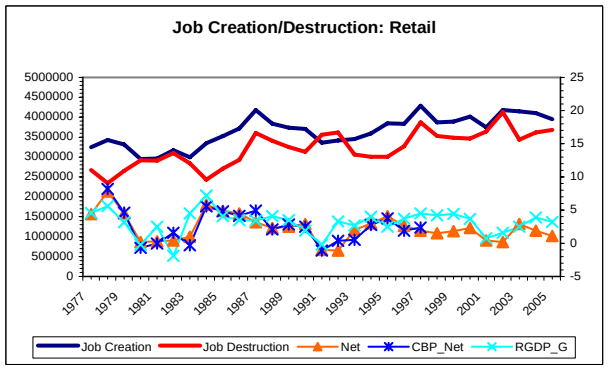
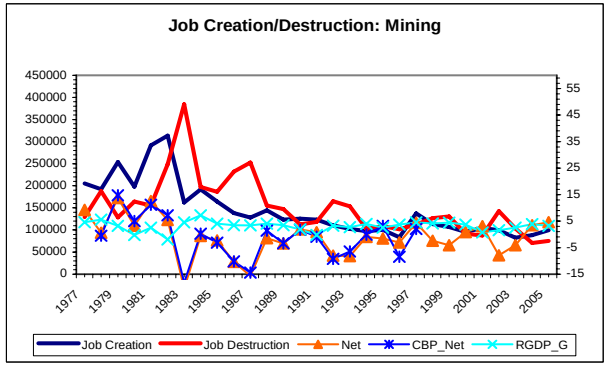


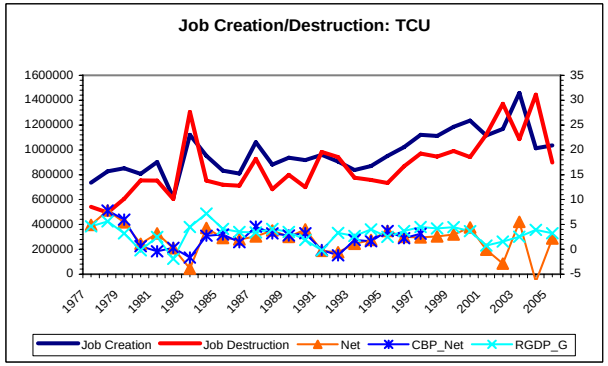
Figure 10

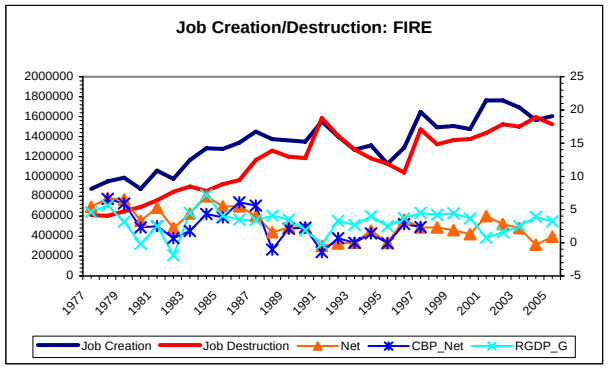

Job Creation/Destruction: Agriculture
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G

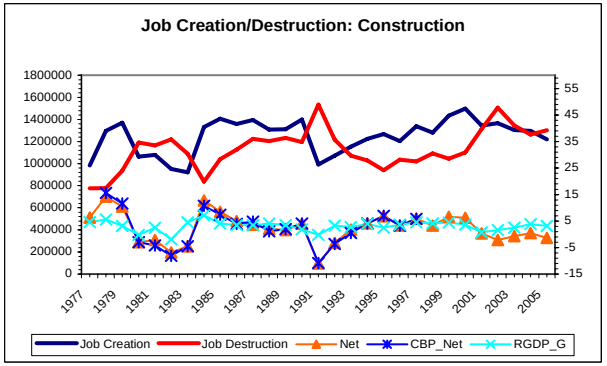

Job Creation/Destruction: Manufacturing
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G

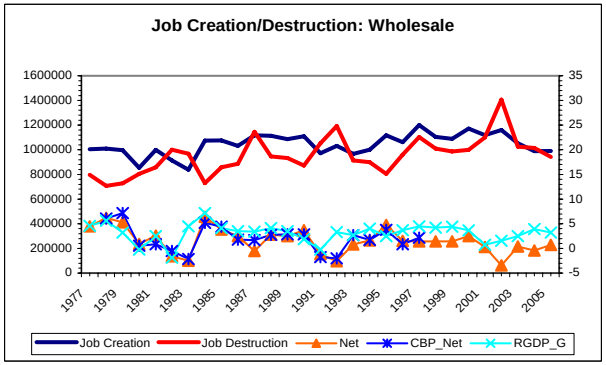

Job Creation/Destruction: Retail
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G


Job Creation/Destruction: Mining
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G


Job Creation/Destruction: TCU
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G


Job Creation/Destruction: FIRE
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G


Job Creation/Destruction: Construction
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G


Job Creation/Destruction: Wholesale
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G

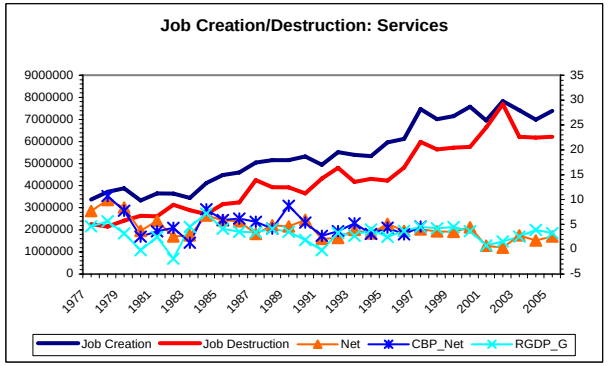

Job Creation/Destruction: Services
Job Creation
Job Destruction
Net
CBP_Net
RGDP_G

Figure 11

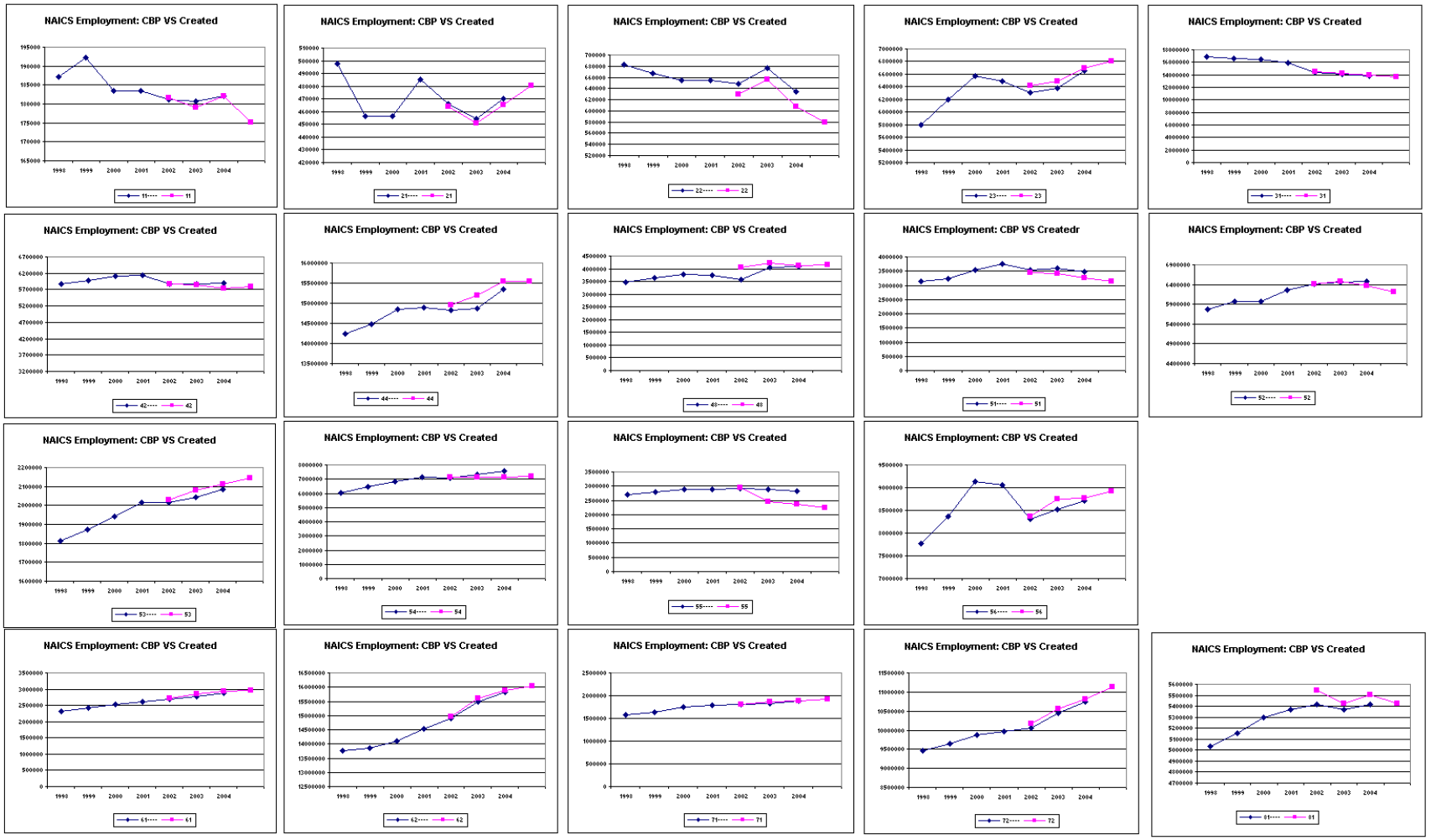

NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Createdr
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created
NAICS Employment: CBP VS Created

Figure 12

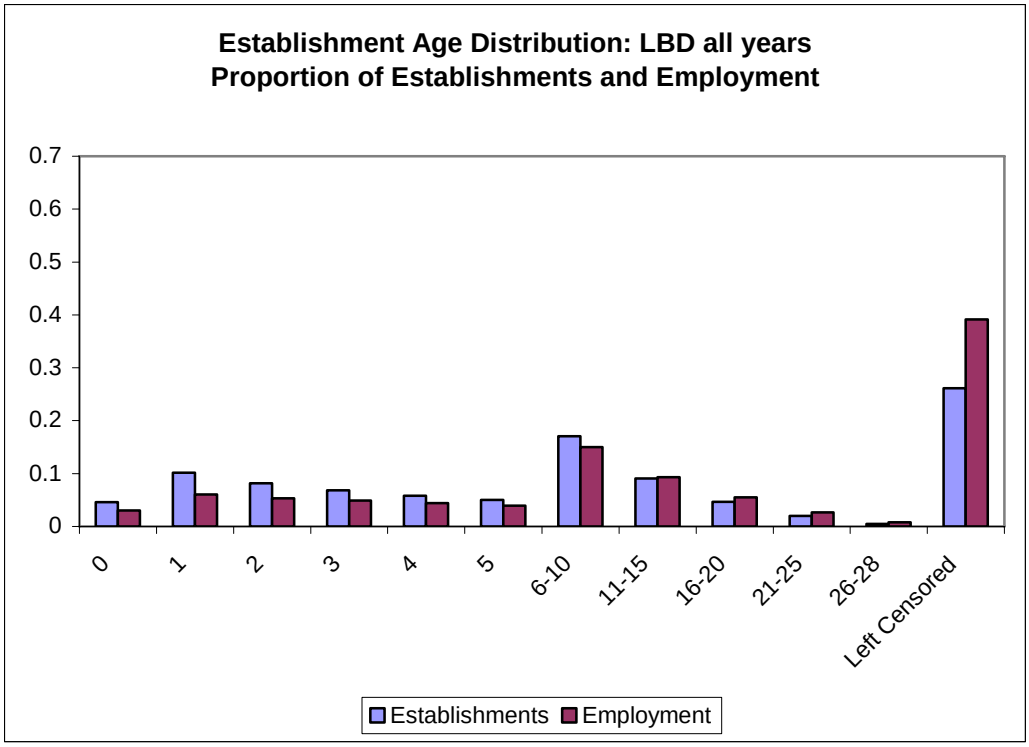


Figure 13

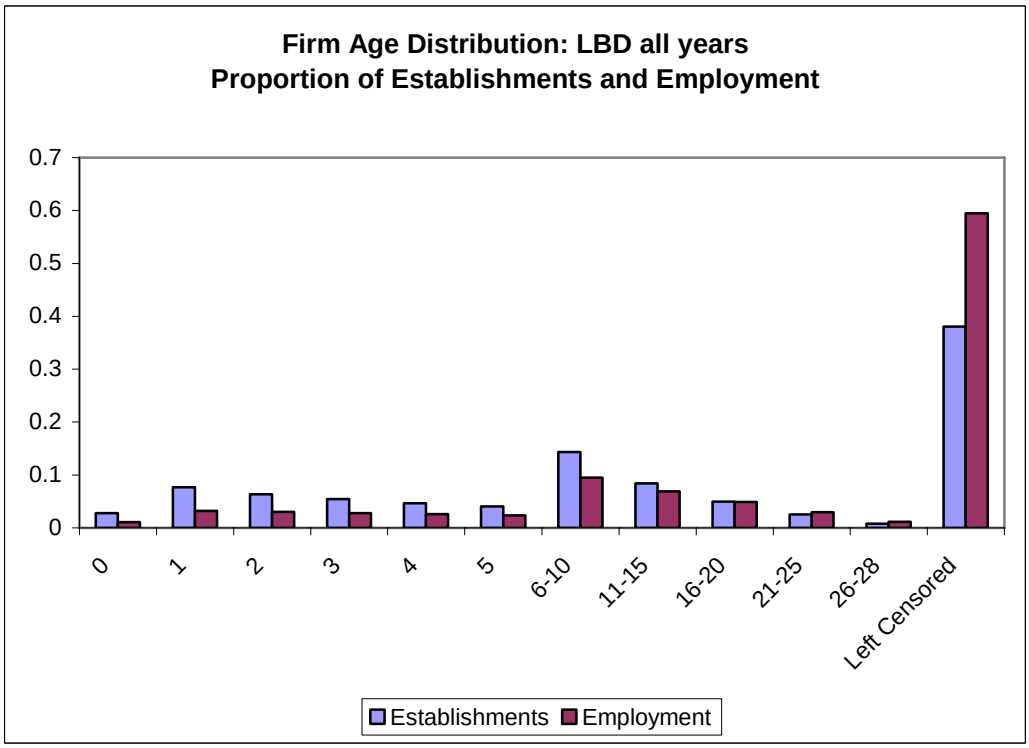


Figure 14