**5/27/2005**


# Births and Deaths in Business Surveys

**Patricia Getz, Jurgen Kropf, Richard Clayton**
**Bureau of Labor Statistics**


**and**


**Ruth E. Detlefsen, Paul S. Hanczaryk, Ron Jarmin, Edward D. Walker**
**U.S. Census Bureau**

# Births and Deaths in Business Surveys

## I.        Introduction

This paper focuses on approaches used to capture births and deaths in business surveys conducted by the Bureau of Labor Statistics (BLS) and the U.S. Census Bureau. Treatment of births and deaths by the BLS will be illustrated primarily through a discussion of the Current Employment Statistics (CES) program in section II. Treatment of births and deaths by the Census Bureau will be illustrated in section III through a discussion of the business register; the current retail, wholesale, and services surveys; and the research work being done at the Center for Economic Studies.  Section IV contains additional information about the BLS Quarterly Census of Employment and Wages (QCEW) which is the BLS sample frame and source of historical business birth/death information used by the CES program.  Section V summarizes the primary purposes, the identification of births and deaths, and the methods used to treat births and deaths for these programs. It concludes by enumerating issues common to all of them.

## II.       Treatment of Births and Deaths in the Current Employment Statistics Program

### Background

The Bureau of Labor Statistics' Current Employment Statistics (CES) program, also known as the payroll survey or the establishment survey, produces monthly estimates of nonfarm employment, hours, and earnings in considerable industry and geographic detail.  The active CES sample includes approximately 400,000 worksites, covering about one-third of total nonfarm employment.  The first preliminary monthly estimates of total nonfarm employment change are published just three weeks after the reference period, making the CES one of the most timely and sensitive national economic indicators.

The CES data are a major economic indicator in their own right, and they also are input to many other key economic series including the national income and product accounts, productivity measures, and indexes of industrial production.

### The CES Sample Design

The CES survey design is a simple random sample stratified by state, industry and size.  The sampling rates for each stratum are defined through an optimum allocation for each state which has a fixed sample size; it is therefore a state-based design. The total nonfarm employment level is the primary estimate of interest and the CES sample design gives top priority to measuring it as precisely as possible, in other words, minimizing the variance at the statewide total nonfarm estimate.  The current CES probability-based sample design was phased-in beginning in 2000; it replaced an outdated quota design that had been in use since the survey's inception in the 1930s. The sample redesign project also provided revised estimators, including net business birth/death modeling to complement the sample-based estimate for continuing business units.

The BLS Quarterly Census of Employment and Wages program (QCEW) provides the micro level files that are used as the CES sample frame, and as input for the modeling of net birth/death employment for the CES, through its Longitudinal Database (LDB). The LDB is derived from State Unemployment Insurance (UI) tax records and contains nearly 8 million individual records which are longitudinally linked back to 1990. The LDB records contain information on monthly employment and quarterly wages covered by the UI system; UI-covered employment comprises just over 97% of in-scope CES employment. The LDB is updated quarterly with information on new business births and deaths. A recent study by BLS confirmed the timeliness of business birth incorporation into the QCEW ("Assessing the Timeliness of Business Births in BLS Establishment Statistics", http://www.bls.gov/ceseta581.study.pdf). A more complete description of the QCEW program is contained in Section IV.

The basic sampling unit for CES is the Unemployment Insurance (UI) tax account; a UI account may have one or multiple worksites. The sampling frame and the sample itself are updated twice a year with new quarters of UI-based universe data. The process helps to keep the sample up to date by adding new firm births and deleting deaths. In addition to the semi-annual update for business births, the CES sample design also provides for a full annual update process which includes sample frame maintenance and the redrawing of the entire sample based on the first quarter of UI data. Frame maintenance provides for the updating of industry, size, and geographic designations. A high degree of overlap (typically about 80%) occurs with each redraw. This results because all UI accounts on the sampling frame are ordered by permanent random numbers (PRNs). The allocation for each sampling cell is fulfilled by working down the ordered PRN list until the full complement of needed units is drawn.

While the sample is updated twice a year, new sample units are introduced into the CES monthly estimation process just once a year, in conjunction with the program's annual benchmark processing. This timing of the new sample introduction is for operational efficiency.

**CES Estimation Procedures**

The monthly CES employment estimates are produced using a two-part estimator: a sample-based component plus a model-based component. The sample-based component of the estimator is designed to measure the employment change from the continuing unit population. Continuing units are defined as those worksites that are in existence on the sampling frame at the time the sample is selected and continue in business through the next annual sample update.

While the CES has a large and representative sample, the sample alone can not measure the contribution of new business birth employment in time for the monthly CES publication, because there is an unavoidable lag between a firm opening for business and its appearing on the UI frame and being available for sampling. There is also some difficulty in estimating business deaths from the sample as it is often problematic to distinguish between a business death and a survey refusal on a real time basis. Therefore CES uses a model-based method to estimate the net contribution of business births and deaths each month. This net birth/death estimate is added to the sample-based estimate to derive the overall CES total nonfarm employment estimate each month. An advantage of modeling the net of births and deaths as opposed to modeling each piece separately or attempting to model only the birth piece, is that it helps negate any linkage problems

2

in the Longitudinal Data Base (LDB).  If a firm changes UI numbers for example and is not identified on the LDB as an actual continuous unit, its effect is cancelled out in the net birth/death modeling as it will count as both a birth and a death and the net effect is zero, as it should be.

In addition to the two-part estimator used to produce current monthly estimates, CES employment estimation also includes an annual benchmarking procedure.  Tabulations from the same UI tax records that are used as a sample frame are used to establish an annual benchmark, or universe count of employment.   The difference between the CES sample-based estimate and the UI-based benchmark for March of each year serves as a measure of how well the CES sample plus model technique has measured the universe employment level and over-the-year change in employment.

**Birth/Death Estimation Technique**

The CES net birth/death estimation technique is based from research which indicated that while both business births and deaths are generally important contributors to employment levels, their net contribution is relatively small and stable, and therefore amenable to model-based estimation. Figure 1 illustrates this with data from the BLS Business Employment Dynamics (BED) program. BED breaks employment change into four components: expansions, contractions, openings and closings at the establishment or worksite level.  The openings and closings are analogous to the births and deaths CES needs to estimate on a monthly basis to derive the total nonfarm payroll employment level.

**Figure 1.**
**Components of private sector gross job gains and losses, seasonally adjusted**
**September 1992 - June 2004**
**Total private**
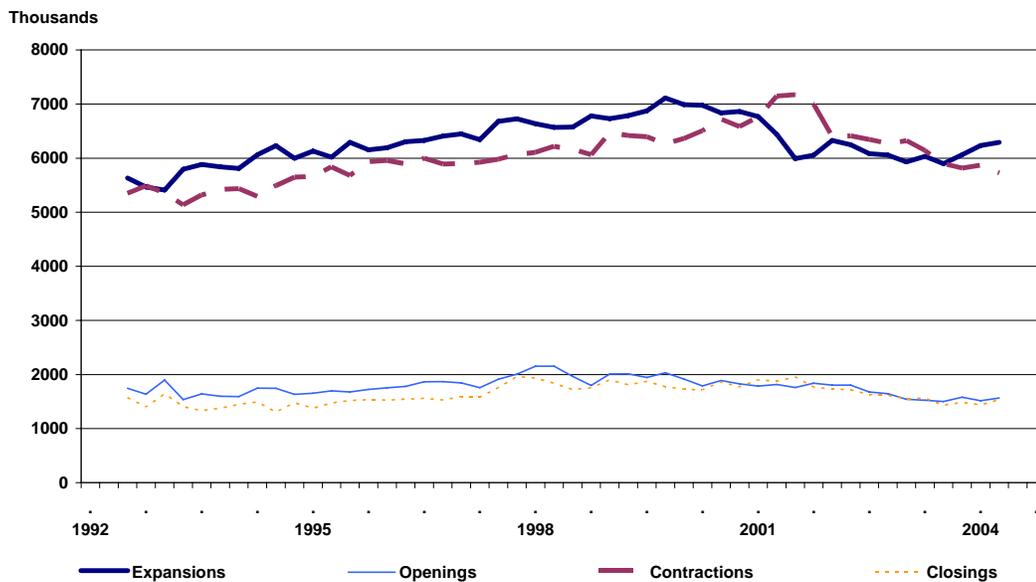


The current net birth/death model was implemented with the CES sample redesign.  It replaced a bias adjustment model previously used in the CES estimates.  The bias adjustment model was

3

used with the old CES quota sample design and intended to correct for limitations in the quota sample design as well as business birth/death employment contributions. As the quota design was phased out and a new probability–based sample phased in over the 2000-2003 time period, the net birth/death model replaced bias adjustment. The combination of the new probability-based sample and net birth/death models performed well in research testing and has performed well to date in production, as measured by the size of the CES annual benchmark revisions, more fully described below.

**Research into Birth/Death Measurement**

Initial research for the probability sample redesign methodology focused on collecting timely information on new UI account registrations, to construct monthly 'real time' birth sample frames. This proved operationally impossible as these early UI tabulations did not match well to data delivered on the full quarterly UI files several weeks or months later.

The second stage of research focused on disaggregating the overall net employment changes into the components derivable from continuing business units, births, and deaths in order to understand how much would be missing from the sample-based estimates and how to best account for it. For any given month the CES employment can be described as:

Total employment = continuing units + births – deaths

It was possible to segment the full universe count of employment from the UI tax records into these categories and study their properties. A major finding from this work was that the birth and death components are each large but their net is small, and stable over time. There also appeared to be some pro-cyclical element to the net birth/death trends, in other words the net birth/death employment grew when the employment from continuing units was growing and declined when employment from the continuing units was declining. These two factors lead to the conclusion that a simple regression or ARIMA model, in conjunction with information provided by the sample of continuing units, might be sufficient for estimating the net birth/death employment for CES purposes. Testing of various model iterations confirmed this.

**Overview of the CES Birth/Death Methodology**

The methodology currently in use for estimating the business birth death contribution in CES is a two part procedure:

*Step 1 – Using sample deaths to impute for unsampled births.* When a firm's employment goes to zero the report is not used as such. The movement of all other sample firms in the estimating cells is attributed to the death firm, in order to partially account for birth firms that are not picked up by the sample. This adds a degree of cyclical sensitivity; if employment in a cell is growing, more birth employment is being imputed. If employment is declining, less birth employment is being imputed.

*Step 2 – ARIMA models for the residual net birth/death* – that part not accounted for by the imputation procedure.

First a birth/death residual series is derived from the UI universe, for each month:

Birth/death residual = Total employment – simulated sample-based estimates

The simulated sample-based series is developed by using the imputation procedure described in step 1 above with the actual UI micro data records to determine how much net birth/death would be missed by estimating from the sample. Next ARIMA models are fit to project these birth/death residual series for the two years following the most recent benchmark. Residual series are developed for each estimation cell for each month; five years of residuals are input to the forecast. Seven common ARIMA models are fit; one with the lowest forecast error is selected for each cell. Models are re-fit and projections done annually in conjunction with benchmark revision process.

**Details of the Birth /Death Modeling Technique**

This methodology assumes that the probability sample accurately represents the QCEW universe. By using the universe data as input in the net birth/death model and applying the same methodology to the data as is applied to the sample when accounting for business births, the historical imputation error can be measured with certainty.

The input data from the LDB reflects the different levels of information regarding business births available to the sample over time. The sample is selected from this database in the first quarter of every year with a lag of 12 months. At the time of sample selection all business births, up to the selection quarter, are known and these units are eligible to be selected for the sample. At this time (March of every year) the monthly CES estimates are benchmarked as well and under the redesign methodology a new estimation cycle begins for the post benchmark months. The post benchmark period is defined as the April following the benchmark through the following October. In order for the LDB to reflect these information levels, 25-month frames are created beginning in March of every year; for example, the frame for the benchmark year 2004 starts in March of 2002 and ends in March 2004.

These frames are used as input for the net birth/death model which iterates through the following steps to calculate the imputation error for each cell:

1. The model creates a subset from the universe of continuous establishments (CE), defined as establishments with employment ($EMP_{BMK}$) greater than zero for the benchmark month. These establishments are eligible for selection in the sample.

$$CE \text{ where } EMP_{BMK} > 0 \qquad (1.)$$

2. Each record in this subset is scanned from the first to the last month of the frame for non-zero employment. When a record displays zero employment ($EMP_t$), an out-of-business firm is detected and the employment for this month is imputed ($EMP_{t,imp}$). The imputed value is calculated by applying a link ($LK_t$) to the previous month's employment and replacing the zero value with the calculated value for this month. The

5

link is derived from the sum of the over-the-month change of every record in the cell (C) which does not have zero employment ($\Sigma$ $EMP_{t,EC}$).

$$EMP_{t,imp} = EMP_{t-1} * LK_t \qquad (2.)$$

where t is the month with $EMP_t = 0$ and

$$LK_t = \Sigma EMP_{t,EC} / \Sigma EMP_{t-1,EC} \qquad (3.)$$

where EC are all other establishments in the cell with $EMP_t > 0$.

This process is continued for every month when an establishment displays zero employment. This process also simulates implicitly the imputation procedure during the estimation process.

3.  The records with continuous employment ($EMP_{t,EC}$) for each month of the frame and the records for which employment is imputed ($EMP_{t,imp,C}$) are summed for each cell. This series represents the estimate ($EST_{t,c}$)if the sample were the universe.

$$EST_{t,C} = \Sigma\ EMP_{t,EC} + \Sigma\ EMP_{t,imp,C} \qquad (4.)$$

4.  The imputation error ($ERR_{imp,C}$) is calculated by subtracting the universe-based employment estimate from the employment of the total population ($EMP_{pop,C}$) in the LDB for each cell and month. The total population contains the actual business births and deaths for the time period each frame covers.

$$ERR_{t,imp,C} = EMP_{t,pop,C} - EST_{t,C} \qquad (5.)$$

5.  The resulting imputation error series ($ERR_{imp,C}$) contains 25 months of data for each production cycle from the benchmark month (when the error is zero because the business births are completely captured), to the last month of the cycle, which is March of the second year after the benchmark year. The first year of the cycle represents the post-benchmark year and the second year is the current production year. The cumulative imputation error for the production cycle is generally larger than the cumulative imputation error from the post-benchmark cycle because of the increased time difference from the benchmark month. Therefore, the forecast from the net birth/death model has to be based on two different time spans, the post-benchmark months ($ERR_{PB,C,BMK}$) and the production months ($ERR_{PR,C,BMK}$). The frames of 25 months are split into 12 post-benchmark months and 13 production months.

$$ERR_{imp,C,BMK} =$$
$$\{\ ERR_{BMK,t,C}\ ,\ ERR_{BMK,t+1,C}\ ,\ \ldots,\ ERR_{BMK,t+12,C}\ ,\ \ldots,\ ERR_{BMK,t+24,C}\ \} \qquad (6.)$$

For the post-benchmark forecast:
$$ERR_{PB,C,BMK} =$$
$$\{ERR_{BMK,t+1,C},\ldots,ERR_{BMK,t+12,C}\} \qquad (7.)$$

For the production forecast:
$$ERR_{PR,C,BMK} =$$
$$\{ ERR_{BMK,t+12,C} ,\ldots, ERR_{BMK,t+24,C} \} \qquad (8.)$$

This procedure is repeated for each benchmark year and the corresponding sections from each frame are joined to form a time series, which consists of over-the-month changes of the errors, as input for the net birth/death forecast ($NETBD_{PB/R,C}$). The input series consists of 5 benchmark years. This example shows the inputs for the benchmark-2004 cycle.

For post-benchmark forecast:
$$NETBD_{PB04,C} =$$
$$ERR_{PB,C,BMK00} \,\|\, ERR_{PB,C,BMK01} \,\|\, ERR_{PB,C,BMK02} \,\|\, ERR_{PB,C,BMK03} \,\|\, ERR_{PB,C,BMK04} \qquad (9.)$$

For production forecast:
$$NETBD_{PR05,C} =$$
$$ERR_{PR,C,BMK00} \,\|\, ERR_{PR,C,BMK01} \,\|\, ERR_{PR,C,BMK02} \,\|\, ERR_{PR,C,BMK03} \,\|\, ERR_{PR,C,BMK04} \qquad (10.)$$

6.   For each cell, two series are forecasted from the input series, $NETBD_{PB04,C}$ (for the post-benchmark period) and $NETBD_{PR05,C}$ (for the production period.)  The forecasts are generated by the standard statistical software program X12-ARIMA which also tests the series for significant outliers.

**Performance of the net birth/death model**

Figure 2 shows the forecasted vs. the actual net birth/death data for the benchmark 2002 production cycle. The actual data is derived from the LDB records made available in the first quarter of 2004.  Table 1 lists the net birth/death amounts applied to the monthly estimates during the post-benchmark period and during the production cycle.

7

**Figure 2.**
**Performance of ARIMA model versus the actual residual for private sector employment**
**April 2002 to March 2004**
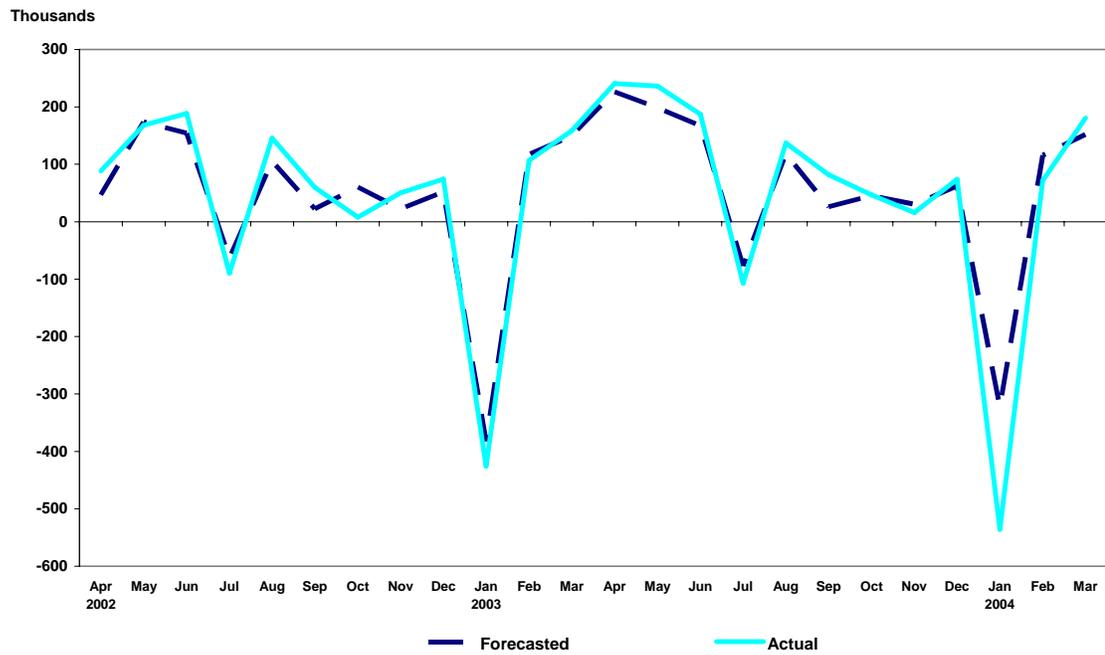
Figure 3 shows the growth rate from the net birth/death model and compares it to the growth rates of the CES estimates without the model and the total CES growth rate for a production cycle.

## Figure 3

| [in thousands) | Growth in employment, March 2002 - March 2003 | | Growth in employment, March 2003 - March 2004 | | Total growth | |
|---|---|---|---|---|---|---|
| | Net change [1] | Rate [2] | Net change | Rate | Net Change | Rate |
| Net birth/death | 150 | 0.14 | 2 | 0.00 | 152 | 0.14 |
| CES estimate, w/o net birth/death | -658 | -0.61 | 518 | 0.48 | -140 | -0.13 |
| CES estimate, total | -509 | -0.47 | 520 | 0.49 | 11 | 0.01 |

**Accumulated growth rate of net birth/death versus estimated CES employment**
**March 2002 to March 2004** [2]

Percent



Legend: ⋯⋯ Net birth/death    ▪ ▪ ▪ Published CES w/o net birth/death    ——— Published CES
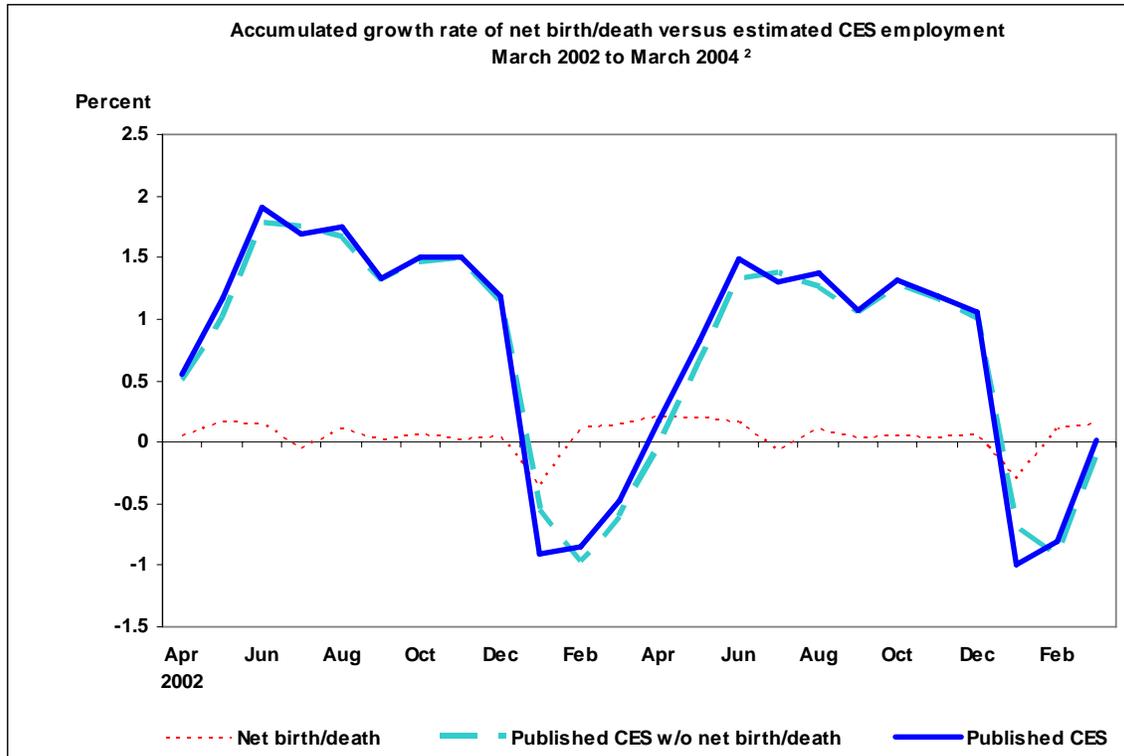
[1] Net change is the sum of employment gains and losses over the time period
[2] The rates measure sum of employment gains and losses as a percentage of the beginning employment level

**Table 1**
**Net birth/death adjustment for total private, April 2003 to March 2005 (in thousands)**

| Year and Month | Total private | |
|---|---|---|
| | Net birth/death adjustment for the post-benchmark period | |
| | 2003 Benchmark | 2004 Benchmark |
| | Monthly Amount | |
| **2003** | | |
| April | 124 | N/A |
| May | 182 | N/A |
| June | 164 | N/A |
| July | -71 | N/A |
| August | 122 | N/A |
| September | 32 | N/A |
| October | 45 | N/A |
| November | 30 | N/A |
| December | 62 | N/A |
| **2004** | | |
| January | -323 | N/A |
| February | 116 | N/A |
| March | 152 | N/A |
| **Cumulative Total** | 636 | N/A |
| (4/2003 to 3/2004) | | |
| **2004** | | |
| April | 270 | 224 |
| May | 196 | 206 |
| June | 183 | 183 |
| July | -92 | -80 |
| August | 119 | 123 |
| September | 39 | 44 |
| October | 44 | 57 |
| November | 53 | 7 |
| December | 78 | 67 |
| **2005** | | |
| January | -387 | -281 |
| February | 119 | 100 |
| March | 161 | 180 |
| **Cumulative Total** | 784 | 827 |
| (4/2004 to 3/2005) | | |

**Benchmark Revisions**

The annual benchmark revision to the CES estimates is commonly viewed as a measure of overall error in the CES estimates. This is not entirely accurate because while the UI-based benchmark data contain no sampling error, they are subject to the same non-sampling error sources common to CES and other surveys, including coverage, response, and processing errors. Nonetheless the benchmark revision does provide a more comprehensive error measure than sampling error alone, and thereby provides a good indication of the overall success of the CES sample-plus-model estimation technique for measuring total nonfarm employment. Since the inception of the CES redesign which introduced both probability sampling and net birth/death modeling, benchmark revisions have generally been quite small, an indication that current estimation techniques are working well overall. The CES program began phasing in the current two part estimator effective after the 2000 benchmark revision. Benchmark revisions for the four years before and since the redesign phase-in began are shown on Table 1 below.

**Table 2**
**CES Total Nonfarm Benchmark Revisions**

| Year | Benchmark Revision | Percent Benchmark Revision |
|---|---|---|
| Pre Redesign | | |
| 1997 | 431,000 | 0.4% |
| 1998 | 44,000 | 0.05% |
| 1999 | 258,000 | 0.2% |
| 2000 | 468,000 | 0.4% |
| | | |
| 4-year mean | 300,000 | 0.3% |
| 4-year mean absolute | 300,000 | 0.3% |
| | | |
| | | |
| Post Redesign (new sample phased in over 2000-03) | | |
| 2001 | -123,000 | -0.1% |
| 2002 | -313,000 | -0.2% |
| 2003 | -122,000 | -0.1% |
| 2004 | 203,000 | 0.2% |
| | | |
| 4-year mean | -89,000 | -0.05% |
| 4-year mean absolute | 190,000 | 0.2% |

**Strengths and Weaknesses of Current Birth/Death Models**

*Strengths of the procedure*
- The imputation part of the procedure is directly related to the current sample and is therefore sensitive to employment trend shifts and turning points.
- Data source for net birth/death model is identical to the benchmark data source thereby minimizing the benchmark error introduced by the model.

11

*Weaknesses of the procedure*

- Data source for the net birth/death model has a lag of 12 months which contributes to the forecasting error of the model.
- Imputed values for the current month's estimate cannot be explicitly measured.
- Net birth/death model is a simple forecasting procedure based on previous years' values.
- Errors in data source (LDB) are transferred to the net birth/death forecast.

*Future research to improve procedure*

- Net birth/death model might be improved by introducing additional variables or relating the model to the sample based estimates.
- Use of more timely input data in the net birth/death model might improve its accuracy.

## Summary

The BLS CES program defines births and deaths in a way designed to advance the overall goal of the program: measuring total nonfarm employment levels on a current monthly basis. The focus for CES is not the precise definition or identification of business births and deaths per se, but making sure their contribution to total employment is properly accounted for such that the total nonfarm employment level is accurate. In that sense the CES birth/death definition is unique to the CES program; essentially births and deaths are those new opening and closings that generate employment gain or loss too recent to be captured by the CES sample. CES solves this problem by using a model-based approach to estimating the contribution of business births and deaths to total nonfarm employment. Annually the efficacy of the CES overall sample-plus-model estimation method is evaluated through a comparison of CES published estimates to a benchmark derived from a near universe count of Unemployment Insurance tax records. The CES birth/death estimation approach is relatively new; it was implemented in conjunction with a new CES probability-based sample design over the 2000-2003 period. To date this new CES estimator has performed well as measured by relatively small annual benchmark revisions.

This section has highlighted the treatment of births and deaths for the Current Employment Statistics program at the BLS. The following section shifts its focus to Census Bureau programs.

## III.      Treatment of Business Births and Deaths in the Census Bureau Programs

This section discusses the treatment of business births and deaths at the Census Bureau. The section starts with a description of their treatment in the Business Register since it provides the frame for most of the Census Bureau business statistics programs. The section continues with a discussion of the handling of births and deaths in the surveys of retail, wholesale, and services. The section ends with a discussion of measuring business dynamics.

**A. Process to Identify Births and Deaths in the Census Bureau's Business Register**

The Business Register (BR) is the Census Bureau's master business list. The BR is compiled and maintained on a continuous basis. The BR contains basic economic information for over 7.5 million employer businesses and over 21 million nonemployer businesses. The BR serves as the primary sampling frame for the economic census and many of the economic surveys. The BR identifies the universe of active businesses, including up-to-date information on business births and deaths. The Census Bureau obtains information for the BR through business surveys and Federal administrative records.

**Business Register Data Sources**

Every 5 years, the Census Bureau conducts an economic census, collecting data for reference years ending in "2" and "7". In December of 2002, economic census forms were mailed to over 5 million establishments with paid employees. All enterprises that own and operate more than one establishment (multiunits) were mailed a census report form for each location. In total, 180,000 multiunit enterprises encompassing 1.7 million business locations received individual census forms. In addition, economic census forms were mailed to about 3.5 million single-location businesses with paid employees, mostly larger ones. Information for the remaining 2-plus million smaller single-location businesses with paid employees come from Federal administrative records.

The economic census plays a critical role in updating and maintaining the organizational linkages within multiunit enterprises. It provides current information for all multi-location enterprises as well as identifying those single-location employer enterprises that have opened or acquired additional locations (i.e., become "multiunits") since the previous economic census.

Between economic censuses, the Company Organization Survey (COS) is the primary vehicle for maintaining the organizational structure of the larger, multi-establishment enterprises. A sample of approximately 50,000 of the largest multiunit enterprises are surveyed annually. In addition to data inquiries, the enterprises are asked to report operational status for each location, report any new locations, closings, spin-offs, or acquisitions. For spin-offs and acquisitions, the enterprise is asked to report the name and address of the company or companies involved in the acquisition or spin-off. The monthly, quarterly, and annual economic surveys are another source of organizational change information and will be discussed in Section B.

For years other than economic census years, the administrative record information, in conjunction with the COS, are the primary inputs to the BR. The Census Bureau makes extensive use of administrative records from the Internal Revenue Service (IRS), Social Security Administration (SSA), and the Bureau of Labor Statistics (BLS), to construct and maintain the BR. In particular, the IRS provides name and address, first quarter employment, four quarters of payroll, and sales/receipts/revenue data for the millions of single-location employer businesses on the register as well as the same data for the EIN payroll tax entities associated with all multi-establishment

enterprises. These data offer an efficient, low-cost alternative to the direct mail canvass of small employer businesses.

Additionally, administrative records are the BR's exclusive source of information about businesses without paid employees—nonemployers. Generally these are very small enterprises operated entirely by working owners, whose personal income is a withdrawal from profits rather than wages. The nonemployer part of the BR is constructed by identifying business income tax returns for which there is no corresponding payroll tax return (see below for more information on business income tax returns and payroll tax returns) to show that the enterprise has no paid employees. Data for 2002, the most recent year available, show that there were 17.6 million establishments without paid employees, and they had receipts amounting to $770 billion.

The IRS data provide a degree of coverage and quality that would be difficult and too expensive to duplicate through other means. Moreover, a major benefit of the IRS data is its timeliness. The IRS provides us with business data on a weekly basis. The Census Bureau receives information on both new and existing businesses almost immediately upon being filed by the taxpayer. This timeliness is essential to the processing of the current surveys. The major administrative record inputs to the BR are as follows:

IRS—The Internal Revenue Code (Title 26, United States Code), authorizes the IRS to disclose tax information to the Census Bureau "for the purpose of, but only to the extent necessary in, the structuring of censuses…and conducting related statistical activities authorized by law." The Census Bureau obtains business information from these IRS sources:

- Business Master File Entity/Directory (BMF)—The IRS activates a BMF account when a business applies for an EIN (Form SS-4, Application for an Employer Identification Number, is provided in Attachment A). The BMF identifies EIN entities representing all business, organizational, and agricultural taxpayers. The Census Bureau receives an annual BMF extract in May, with a volume of approximately 24 million records. Further, the Bureau receives monthly BMF updates for recent additions, deletions, and other changes, with a total volume of 18 million records. The BMF provides the EIN and other identifying information; legal and trade names; mailing and physical location addresses; industrial classification codes; and selected status and processing indicators. BMF information is critical to the BR, particularly for identifying newly established EIN entities that represent business *births*.

- Payroll Tax Returns—Business employers file the Employer's Quarterly Federal Tax Return, IRS Form 941 to report employment taxes. The Census Bureau receives weekly Form 941 files from IRS processing, with total volume amounting to 23 million records (about 5.8 million Form 941 records per quarter). The quarterly Form 941 content includes the EIN, March 12th employment (for first quarter returns) and three measures of wages: total compensation, social security wages and tips, and Medicare wages. In addition, the Form 941 includes a checkbox for those businesses filing their final tax

return; this oftentimes signifies a business death. The Census Bureau receives timely information from quarterly payroll tax returns, the IRS Form 941 series. The deadline for filing this return is one month after the end of the reference quarter, and the Census Bureau receives over 95% of the quarter's records within three months after the end of the reference quarter. Most of the Census Bureau's surveys cover employer businesses only, and presence of IRS Form 941 payroll defines the BR's business employer universe.

- Business Income Tax Returns—Businesses report their income and taxes due on various form types, depending on legal form of organization (Form 1120 series for corporations; Form 1065 for partnerships; Form 990 series for certain exempt organizations; and Form 1040, Schedule C for individuals who operate sole proprietorships businesses). The annual income tax files provide the sales/receipts/revenues, industrial activity codes, assets, and much more; further, the file for Form 1040, Schedule C provides identifying information and address, as these taxpayers are not included in the BMF. A business income tax return that cannot be associated with an employer enterprise (i.e., a business that also files a payroll tax return) usually is regarded as a nonemployer. [1]

SSA—New business and organizational taxpayers (i.e., *births)* file an Application for Employer Identification Number, Form SS-4, with the IRS. The IRS forwards the file of records to the SSA, which assigns an industrial classification to the applicant based on a principal activity description and other information reported on the form. The SSA provides Census with an EIN, a six-digit North American Industry Classification System (NAICS) code, estimated employment, and other classification/status indicators. The Census Bureau receives monthly files from current SSA processing, which lags Form SS-4 filing by about one to two months. This timing is down significantly from the eight to ten month lag in past years. The SSA has implemented an automated coding program that codes approximately 60% of the SS-4 records. The SSA clerical coding staff assign NAICS codes to the remaining SS-4 records. Approximately 2.4 million Form SS-4 records are obtained from SSA each year.

BLS—The BLS maintains a separate business register, known as the Longitudinal Data Base (LDB), based on information collected in connection with unemployment insurance program administration. State employment security agencies supply industrial classifications and other data about private- and public-sector employers to the BLS as part of the Quarterly Census of Employment and Wages (QCEW) program. Under authority of an Office of Management and Budget (OMB) directive and a memorandum of understanding between the agencies, the BLS shares industrial classifications with the Census Bureau. Each quarter, the Census Bureau prepares a file of EINs that identify unclassified single-units and partially classified single-units

---

[1] Certain other conditions apply. For example, the income tax return of a business entity classified as a mutual fund is excluded from the universe of businesses without paid employees. Also, businesses that file income tax returns showing very large receipts are excluded on the premise that their size makes it highly probable that they are part of an employer enterprise, even though the information available from administrative records does not make the connection apparent.

from the BR. The BLS refers each of these EINs to their Business Establishment List (BEL) and returns the corresponding six-digit NAICS code when available.

## Identifying BR Births and Deaths

The Census Bureau relies on IRS data in the BR to identify employer enterprises. When a new business payroll record is received from IRS (on Form 941), a business employer record is added to the BR. The IRS data are provided at the EIN entity level. The IRS data are extremely timely – the Business Register identifies the majority of employer births within 3 months after the business initially pays wages.

Similarly, the BR relies on business income tax returns to identify nonemployer businesses. When a new income tax record is received from one of the IRS forms used to report business income *and* when that income tax record does not have a corresponding payroll tax record, a business nonemployer is added to the BR (see above for exceptions). Since the business income tax returns are filed annually rather than quarterly, identification may be somewhat less timely. Still, new nonemployers generally are added to the BR no more than a few months after the end of their annual tax reporting period.

The Census Bureau identifies new establishments of multiunit firms through Census Bureau mailings, and primarily through the 5-year Economic Census and the annual Company Organization Survey (COS). The companies provide information on establishment openings and closings. The annual COS surveys approximately 50,000 enterprises: 45,000 of the largest multiunit enterprises, plus another 5,000 or so that have shown significant changes in its IRS employment or payroll data. Significant changes in the affiliated IRS data of a multiunit enterprise typically alert us of birth or death activity.

Further, the Census Bureau uses IRS Form 941 data to identify single-unit business deaths. A code on IRS Form 941 indicates when the employer will no longer use the EIN for tax reporting purposes. When this Form 941 filing requirement code is set to inactive, and the stream of quarterly Form 941 payroll tax reports cease, the business will be removed from the current surveys. Similarly, the death of a nonemployer is signaled by a cessation of business income tax reporting. Multiunit establishment deaths are identified through information collected in the annual COS and the 5-year Economic Census.

## Issues Relating to BR Births and Deaths

Although new activations can be identified from administrative record sources (as indicated by new EINs), these new entities do not necessarily represent true births. Many represent legal reorganizations. As an example, many sole proprietorships become corporations, which requires a new legal entity represented by a unique EIN. At the BR level, the Census Bureau has not developed a methodology to link these and other types of business reorganizations across years,

except through the quinquennial economic census collections that obtain information about a continuing business's new EIN.

The Longitudinal Business Database (LBD), maintained by the Center for Economic Studies (CES) has filled this void.  The LBD links establishment records across years, which permits CES to produce meaningful birth/death statistics.  The CES work on births and deaths, and the implications for their analytical work is covered in Section C.

**B. Births and Deaths in Surveys of Retail, Wholesale, and Services**

<u>Survey Details</u>

The Census Bureau conducts monthly and annual retail and wholesale and quarterly and annual services sample surveys. These surveys collect data used to compute estimates of dollar volume and percent change at the U.S. level for various kinds of business as defined by the latest industrial classification standard recognized by the Office of Management and Budget. Data collected for the retail and wholesale sectors include sales, inventory, and value of purchases. Data collected for the services sectors include receipts and various expense and product data for particular industries.

Approximately once every five years, soon after the conduct of the most recent economic census, new samples representing the employer universe are selected for these surveys.  Nonemployers are represented directly from administrative records in the annual retail and services surveys and through a benchmarking procedure for the monthly retail survey. The procedure benchmarks the monthly survey sales estimates to the latest annual survey results.

Sampling units for employers consist of companies or Employer Identification Numbers (EINs). Companies are multi-establishment (MUs) businesses. EINs, issued to businesses by the IRS, may be either single establishments (SUs) or MUs. Stratified random samples are selected for each of the sectors. Company units are always selected with certainty while EINs may be selected with certainty or with weight greater than one.

An integral part of maintaining the quality of the estimates produced from these surveys is assuring that the samples represent the target population for the reference period of the estimates. Therefore it is necessary to update the samples regularly to reflect ongoing changes to the universe.

<u>Sample Maintenance</u>

Samples are maintained through procedures designed to identify and update for EIN births, reactivations, and deaths; spin-offs, splits, mergers, acquisitions for certainty units; and openings or closings of establishments within a sampling unit.  Administrative data from the Business Register (BR) is used to identify EINs that are births, reactivations, or deaths.  Other changes to

the universe are identified through such things as information collected on survey reports forms, follow-up contacts with companies, publicly available sources, and annual matches to the BR. EIN births, reactivations, and deaths are discussed first, followed by spin-offs and splits, then mergers and acquisitions, and finally openings or closings of establishments within a sampling unit. The discussion that follows is for the monthly and annual surveys, but not for the quarterly survey. Methods for the quarterly are similar, though not exact to those used for the monthly surveys.

## EIN Births, Reactivations, and Deaths

Administrative data from the BR are used to determine EIN births and reactivations, and to determine EIN deaths for the noncertainty portion of the samples. Other methods are used to identify deaths for the certainty EINs and companies.

EIN Sampling Units

Samples are periodically updated to represent EINs issued by the IRS since the initial sample selection. These new EINs, called births, are on the latest available IRS mailing list for Federal Insurance Contribution Act (FICA) payroll tax payers. They may have a kind-of-business classification that was assigned by the IRS, SSA, or the BLS.

EIN births are subjected to sampling on a quarterly basis using a two phase procedure. To be eligible for subjection to first phase sampling, a birth must satisfy these conditions:

- was not previously subjected to sampling or identified as out-of-scope
- is classified in-scope to the retail, wholesale or services sectors
- if not classified, has more than 3 employees or nonzero payroll in one or more of the latest four quarters
- is Business Master File (BMF) active, i.e., has an active filing requirement for IRS Form 941, Employer's Quarterly Federal Tax Return
- has a valid state code.

In the first phase, births are stratified by kind-of-business (including none specified) and measure of size (MOS) based on expected employment or quarterly payroll. The MOS actually assigned is the one that gives the birth the best chance of selection. In most instances this is the payroll based MOS since payroll is available more often than employment. Equal probability systematic sampling is done within each stratum. About 125,000 - 150,000 births are identified each quarter from which approximately 10% are selected in the first phase.

Each selected birth is canvassed to obtain additional information (type of operation, company organization information, taxable or tax-exempt status, wholesale inventories, and electronic commerce information) from which to derive a new or more detailed kind-of-business classification. In addition, sales or receipts for two recent months are requested for use in computing the MOS. The survey is not mandatory and the response rate is about 70%. For the

18

births that do not provide the monthly data, the MOS is computed from administrative payroll or employment.

The additional information obtained from this canvass may indicate that some selected first phase births are out of scope or associated with units already represented by the samples. In addition, some births may not provide enough information for further sampling. These are not subjected in the second phase. For the remaining births, the MOS computed from the more recent data is used to determine the second phase sampling weight.

Sampling for the second phase is done three months after that for the first phase. Births are stratified by the kind-of-business and MOS determined from the first phase canvass. Probability proportional to size systematic sampling with overall probabilities equivalent to those used in selecting the initial samples is done within each stratum. About 20% of the first phase selects are selected in second phase sampling.

Selected second phase births are added to the monthly and annual samples. The average amount of lag for a birth to be added is 6-9 months. The lag is influenced by a number of factors including the time between the birth starting operation and filing of the Application for Employer Identification Number (IRS Form SS-4) and the time needed to accomplish the two-phase birth-selection procedure.

If a selected first phase birth does not provide sufficient information for second phase sampling, it is retained on an unsampled birth register. If administrative data updates from the subsequent quarter provide sufficient data, the birth on the unsampled birth register will be subjected to second phase sampling the next quarter. Otherwise, the birth is canvassed again the following quarter if it is still in operation and has recently reported payroll. This every-six-months mailing will continue until a response is received or sufficient administrative data becomes available. For any given quarter there are about 17,000 births on the unsampled birth register. Most of these do not have sufficient industry classification for sampling purposes.

The processing of the EIN births differs between the monthly and annual surveys. For the monthly surveys, births selected in one quarter are added into the survey during the next quarter. Because births are not represented in the monthly sample until they go through the two-phase selection procedure, an interim procedure is used to account for births during the period between the onset of activity and the time of birth selection. This consists of imputing data for all EINs selected in the monthly sample that are known to be out of business, but still have active IRS filing requirements.

Concurrent with the quarterly introduction of births into the monthly surveys, the Census Bureau reviews the IRS payroll tax filing requirement of every EIN selected since the initial sample. Any EIN that was previously inactive and is now active is reactivated and any EIN that was previously active and is now BMF inactive is made a death.

The monthly surveys are updated once each quarter to simultaneously introduce the EIN births, deaths, and reactivations. When the updates for a quarter are first applied to the surveys, sampling weights are adjusted for them to smooth out the effects of introducing all at once, changes that have actually occurred over time. For the first month in which the updates are included in the monthly surveys, data for each birth and reactivation is tabulated using one third of its full weight. For the second month, the data is tabulated using two thirds of its full weight. For the subsequent months, the data is tabulated using its full weight. Imputed data for each death is tabulated using two thirds of its weight for the first month, one third its weight for the second month, and zero weight for subsequent months.

The annual surveys are updated prior to mailing to include all EIN births and reactivations that were not included in the prior annual survey. To better represent all EIN births in the annual survey year, and to account for the lag in processing EINs actually born during the reference year, the Bureau also adds births that were selected in the first three quarterly operations of the year following the survey reference period. Data for these EINs are obtained through supplemental mailings. If an EIN was BMF inactive at the end of the annual survey reference year, but was active sometime during that year, the EIN is not mailed, but data is imputed for the period for which the death was active.

Deaths for Certainty EINs and Companies

For both the monthly and annual surveys, EINs selected into the samples with certainty are not automatically dropped from canvass and tabulation when they become BMF inactive. Rather, the firm that used the EIN is contacted, and if an active successor EIN is found, the successor is added to the survey. Only then will the inactive EIN be dropped.

BMF activity is not used to maintain certainty companies.

Spin-Offs and Splits

A company or MU EIN may spin off some of its establishments which become one or more separate legal entities or parts of other existing legal entities. In considering the outcomes of a spin-off or split, note that "old EINs" refers to the original EINs associated with spin-off establishments. Also, "successors" refers to the separate legal entities acquiring the establishments that were spun off. Four possible outcomes of a spin-off or split are:

- All old EINs are voided and new EINs are obtained by the successors.
- One successor retains the old EINs while the other successors obtain new EINs.
- All old EINs are retained and split amongst the successors.
- One or more of the successors retain some old EINs and obtain some new EINS.

If either of the first two occurs, the new EINs will eventually be subjected to sampling as part of the birth procedures. When a noncertainty EIN splits, these two possibilities are most common. In this case, no special action is required.

When a certainty sampling unit splits, the original sampling unit no longer exists, and at least two new sampling units are created. Each successor unit is added to the sample with certainty. This is done to maintain coverage and eliminate nonsampling error.

When a certainty sampling unit spins off establishments, the original unit remains and the establishments that were spun off could be acquired by units already subjected to sampling, by units not subjected to sampling, or some combination of these. If a large proportion of the original unit was acquired by units in the noncertainty universe, all of the acquiring units are assigned to be certainty. Retaining them as noncertainties, could increase the sampling errors dramatically. Reassigning them as certainties, increases the bias. Thus, the reassignment is a trade off between sampling error and bias. If the bulk of the original unit moves to the nonsampled universe, the spun off components are made certainty units.

<u>Mergers and Acquisitions</u>

Mergers and acquisitions involve reductions in the number of sampling units. A merger occurs when at least two sampling units (the predecessors) combine to form a single sampling unit (the successor) with a different identifier – either a new EIN or a new company. An acquisition occurs when the successor maintains the same identifier as one of the predecessors. The possibilities include:

- All predecessors are certainty units.
- At least one predecessor is certainty and at least one is a selected noncertainty EIN.
- At least one predecessor is certainty and at least one is a nonselected EIN.
- All predecessors are selected as noncertainty EINs.
- At least one predecessor is a selected noncertainty EIN and at least one is a nonselected EIN.
- None of the predecessors are selected into the sample.

If the predecessors are all certainty, they are dropped from the sample and the successor is added as a certainty sampling unit. The same is done for an acquisition.

If at least one predecessor is certainty and at least one is a selected noncertainty EIN, the Bureau drops the predecessors and adds the successors to the sample as a certainty. In doing this, the assumption is made that the noncertainty unit in the sample represents other EINs subjected to sampling but not selected that have been or will soon be acquired by certainty companies.

If at least one predecessor is certainty and at least one is a nonselected EIN, the merged unit is added as a certainty and the predecessors are dropped.

If all predecessors are selected as noncertainty EINs, the sampling units are retained in the sample, however their data is imputed, so that they are maintained in the sample as originally sampled.

If at least one predecessor is a selected noncertainty EIN and at least one is a nonselected EIN, the selected sampling units are retained in the sample, however their data is imputed, so that they are maintained in the sample as originally sampled.

If none of the predecessors are selected into the sample, nothing is done.

Openings and Closings of Establishments Within a Sampling Unit

A company or EIN may open new establishments or close existing establishments. These types of changes are identified during data collection. Data obtained on the report form reflect these changes.

**Effect of Sample Maintenance Operations**

EIN Births, Reactivations, and Deaths

After new samples are selected but prior to surveying them for the first time, the samples are updated with EINs born during the interim. Likewise, the BMF activity is updated for each selected EIN to identify deaths. The new samples selected in January 2000 represented the universe as of July 1999. Table 3 shows counts from the January 2000 sampling operations.

**Table 3**
**Sample Counts from the January 2000 Reselection**

| Survey[2] | Certainty Companies | Certainty EINs | Certainty Total | Noncertainty EINs | Total |
|---|---|---|---|---|---|
|  |  |  |  |  |  |
| MRTS | 2,169 | 689 | 2,858 | 8,776 | 11,634 |
| MTS | 1,078 | 302 | 1,380 | 2,561 | 3,941 |
| ARTS | 2,422 | 782 | 3,204 | 18,926 | 22,130 |
| ATS | 1,078 | 302 | 1,380 | 5,122 | 6,502 |
| SAS | 8,487 | 7,439 | 15,962 | 46,351 | 62,277 |

After being updated for births and deaths, the new samples were mailed for the first time in early 2000. After the initial update, the samples have been updated quarterly for births, reactivations,

---

[2]    MRTS = Monthly Retail Trade Survey;  MTS = Monthly Trade Survey;  ARTS = Annual Retail Trade Survey;  ATS = Annual Trade Survey;  SAS = Services Annual Survey

and deaths. The number of births, reactivations, and deaths to the monthly retail and wholesale samples for each quarter of 2003 are provided in Table 4. The initial number of updates made to the new samples was approximately equal to the number made for two quarters of regular ongoing updates.

**Table 4**
**Number of Births, Reactivations, and Deaths to the Monthly Retail**
**and Wholesale Surveys for 2003**

| | --------------Retail-------------- | | | | ----------------Wholesale--------------- | | | |
|---------|--------|---------------|--------|------|--------|---------------|--------|------|
| Quarter | Births | Reactivations | Deaths | Net | Births | Reactivations | Deaths | Net |
| 1 | 452 | 56 | 325 | 183 | 76 | 12 | 77 | 11 |
| 2 | 425 | 128 | 415 | 138 | 45 | 24 | 83 | -14 |
| 3 | 389 | 41 | 461 | -31 | 60 | 4 | 92 | -28 |
| 4 | 537 | 65 | 431 | 171 | 62 | 21 | 80 | 3 |

For each month of 2003, about .5 - 1.5% of retail sales and .5 - 1.0% of wholesale sales came from the newly identified birth and reactivation EINs.  Births added to the monthly samples during 2003 that were still active in December 2003 contributed 2.0% and 1.2%, respectively to the December sales estimates for retail and wholesale.  Births accounted for 1.9%, .9%, and 1.8% of sales, respectively for the 2003 annual surveys of retail, wholesale, and services.  Because births contribute only a small percentage of total sales for any given month, their impact on the current- to-prior month change estimates will be minimal at the total level. Their impact may be greater for certain industries. The births added over a given year will impact estimates of year-over-year change to a somewhat greater extent. It should be noted, however, that the change in EIN sampling units over the life of a sample is considerable, so that it is important to keep maintaining the samples. For retail, about 43% of the EINs included in the monthly sample just prior to the last sample revision were births added since the sample revision prior to that one.

Spin-Offs, Splits, Mergers, and Acquisitions

During the period between the sample reselection and its initial mailing, the selected certainty units are reviewed to ensure that organizational changes are correctly reflected in the survey. Since most, but not all, of the certainties in the new sample were included as certainties in the prior sample, the number of organizational changes identified will be somewhat higher than at other times in the cycle of the survey.  The approximate numbers of such changes were 50, 12, and 815, respectively for the 2003 annual surveys of retail, wholesale, and services.

**Issues**

The quarterly EIN birth, reactivation, and death process and the procedures for handling spin-offs, splits, mergers, acquisitions, and opening or closing of establishments within a sampling unit, assist in keeping the samples representative of the survey universes.  These were designed to

identify and account for EIN births not already represented by the samples, inactive EINs previously represented by the samples that subsequently reactivate, active EINs represented by the samples that subsequently go out of business, and other organizational changes to EIN and company units represented by the samples.

Some issues of concern include the accuracy of business classifications assigned to the births, the timeliness of the EIN birth processing, and the rigor of methods used to account for spin-offs, splits, mergers, and acquisitions.

Business Classifications for EIN Births

When EIN births become eligible for first phase sampling, business classification may come from the SSA, the IRS, or the BLS, or may be unavailable.  Until recently, about 40% of the births subjected to first phase sampling in any quarter have been unclassified.  With the implementation of an automated coding system at SSA, this percent has dropped to 20 - 30%.  Many of the SSA classifications are only partial.  About 20% of the units mailed out with a particular industry code are recoded to an entirely different industry code (not necessarily in-scope to the surveys).  About 45% of the first stage codes matched to the response code at detail sufficient for sampling. Thus it is necessary to mail a form to obtain accurate and complete classification for a sample of EINs.

Budget permitting, classification could be further improved by expanding the first phase sample to:
- include births currently considered out-of-scope to the surveys (if mailed, some would be found to be in-scope)
- increase the first phase sample to target industries having the most classification issues,
- increase the sampling rate for smaller units, and
- lower the thresholds used to select births as certainty.

In addition to improving classification, expanding the first phase sample would allow us to collect information such as company affiliation and locations of operations from additional EINs. Having this information would improve the coverage of the existing company and MU EIN sampling units.

Timeliness of the EIN Birth Process

Ideally, EIN births should be represented in the samples as soon they have business activity. With the two phase sampling process, they are represented about 6 to 9 months after they first have business activity.  Births are introduced into the monthly surveys once each quarter.  If births were introduced every month the lag would be reduced to 6 months.  The births have not been introduced monthly because of administrative data and response time constraints.

Administrative data used to identify in-scope EIN births are not received in a uniform manner during the year.  The receipt pattern is more uniform on a quarterly than on a monthly basis.  Introducing EIN births monthly could artificially affect the month-to-month change estimates.

EIN births selected in the first phase sampling are given 30 days to respond to the form sent to obtain classification and measure of size information.  Shortening the collection period would require us to gain the approval of the Office of Management and Budget.

The Census Bureau has considered reducing the lag and cutting costs by using one phase sampling.  However, this would require accurate and complete classification and measure of size information from the administrative sources.

Methodology Behind the Spin-Off, Split, Merger, and Acquisition Procedures

Rigorous methodology to account for spin-offs, splits, mergers, and acquisitions would require complex implementation procedures.  Consequently, several simplifying assumptions have been made to streamline the procedures.  The Bureau has tried to strike a balance between precise statistical methodology and the limits of the processing systems.  While the solutions may not be ideal for either, they should be practical for both. Black (2001) presents a more detailed discussion of this.

## C.  Measuring Business Dynamics for Research Purposes

### Background

The Census Bureau's Center for Economic Studies (CES) constructs and maintains longitudinal datasets for use by researchers at CES and a network of Research Data Centers (RDCs).  The first of these datasets was the Longitudinal Research Database (LRD) that contained establishment level data from the Census of Manufacturers and Annual Survey of Manufacturers linked over time.  Work in the 1980's and 1990's with the LRD gave researchers the first solid evidence of the tremendous churning and heterogeneity present in the U.S. economy (see Dunne, Roberts and Samuelson, 1988 and Davis, Haltiwanger and Schuh, 1996).  Given the success of research with the LRD, researchers naturally wanted to see if similar patterns held for non-manufacturing industries.

To meet this need, the Longitudinal Business Database (LBD) was developed at CES (see Jarmin and Miranda, 2002).  The LBD was constructed by linking annual snap shots of the Census Bureau's employer Business Register (BR).  It currently includes establishment level data for all private non-agricultural sectors of economy from 1975 to 2001.  Since the source data for the LBD is the BR, data quality issues on BR are transmitted to the LBD.

One data quality issue that is particularly important to the measurement of establishment births and deaths is the timing of MU births over the 5 year economic census cycle.  During non-census

years, the Census Bureau collects establishment information from only very large firms and a sample of smaller MU companies via the Company Organization Survey (COS).  All MU companies are asked to report their establishment level data in economic census years.  As a result, there is a spike of MU establishment births and deaths in economic census years in the LBD.  CES has nearly completed an algorithm that adjusts these MU establishment births and deaths in the historical data.  However, better BR source data would improve future LBD data sets.  This would most likely require expanding the coverage of the COS or acquiring Multiple Work Site Report data from BLS under the Confidentiality Information Protection and Statistical Efficiency Act (CIPSEA).

The LBD is designed to be a research dataset that serves the varied needs of researchers at the CES and the RDCs.  It contains only basic establishment information derived from the Business Register.  This information includes location, industrial classification, firm ownership, employment and payroll.  However, establishment information in the LBD can easily be linked to other files at the CES and the RDCs that contain detailed data from the economic censuses and other Census Bureau establishment and firm level surveys.

The LBD is used primarily to study business dynamics.  This includes analyses of business entry and exit (births and deaths), firm growth, ownership change and changes in industrial structure.  Linked to more detailed economic census and survey data, the LBD can be used to understand productivity and labor market dynamics.  Since the focus here is on measuring business births and deaths, the paper will now provide a more detailed discussion of how researchers can use the LBD to examine this issue.

## Measuring Entry and Exit with the LBD

The LBD was designed to be used as a research micro data set.  As such, it contains information that allows researchers to compute business entry and exit in a variety of ways that are tailored to the research question they wish to address.  This is important since the theoretical concept of business entry and exit that economists have in mind usually does not correspond to the operational notions of business birth and death used by statistical agencies.

Economic theory is concerned with the operation of markets.  Entry and exit in this context refers to the entry or exit of a particular decision making unit, or firm, into or from a market.  While this definition provides for elegant textbook expositions, applying it to real world situations is much more difficult.  Defining markets in a systematic manner is extremely difficult and far too costly an activity for statistical agencies to engage in.  The Census Bureau's Statistics on U.S. Business programs uses the Business Information Tracking Series (BITS), a production dataset that shares many characteristics with, and is constructed from the same primary data as the LBD, to measure establishment births and deaths by state, sector and enterprise (firm) size.

Figure 4.

**Establishment Birth and Death Rates: by Year**
**Source: LBD**



Death rate — Birth Rate

Figure 4 above shows establishment birth and death rates over the period covered by the LBD. Establishment births and deaths are relatively straightforward to track in the register based BITS and LBD.  Establishments are physical locations.  When a business commences activities at a given location an establishment record with a permanent longitudinal identifier[3] is created in the BR.  This shows up as an establishment birth in the BITS and LBD.  As long as activity continues at that location, the establishment is considered a continuer.  An establishment death occurs when activity ceases.

An additional advantage of the LBD versus other longitudinal files is its longer time series.  This enables us to see, as in the chart above, that the birth and death rates for U.S. employer establishments have been trending down over the last 25 years.

Establishment birth and death statistics produced from the BITS and LBD are very useful but can not address all the questions researchers have about business dynamics. In particular, they do not conform well to the economist's notion of firms entering and exiting markets.  However, the LBD provides researchers additional flexibility in producing estimates of business dynamics.  This

---

3    Note that before the early 1980's the Census Bureau's BR, formerly know as the Standard Statistical Establishment List (SSEL) did not have longitudinal establishment identifiers except for manufacturing plants.  In creating the LBD, CES assigned longitudinal IDs to all establishments in all sectors.

flexibility is possible since the Census Bureau maintains basic information on all business establishments in the BR. Of particular importance is information on the location and industrial classification of establishments, their firm affiliation, and measures of size such as payroll and employment. The LBD then carefully tracks these establishments over time which gives the timing of establishment births and deaths, as well as the ability to measure establishment growth and changes in ownership. Establishments can also be aggregated up to enterprises and enterprise segments.

This seemingly modest amount of information can be used to construct a wide variety of measures of business dynamics that get closer to the economist's preferred measure. In particular, it is possible to more carefully define the scope of the markets that firms are entering and exiting by using some combination of geographic and industrial classifications. For some industries, especially many manufacturing industries, the appropriate market is the entire U.S. (if not the whole world). For others, such as retail and many services, markets are more local in nature. In these cases, it may better to define the market as a particular NAICS code operating within state, county or smaller unit of geography. While these crude groupings undoubtedly have shortcomings, they allow for customization of market definitions across industries.

Once the more accurate market definition is obtained, researchers can measure the entry and exit of firms in these markets. For example, Jarmin, Klimek and Miranda (2005) examine the dynamics of U.S. retail markets over the 1975 to 2000 period. They define markets as a 4 digit Retail Standard Industrial Classification (SIC) industry – County pair. A firm is said to operate in a given market if it has one or more establishments operating within the given SIC – county pair. Firms can enter these local markets in a variety of ways. They can build one or more establishments in the SIC – county pair, they can purchase an establishment in the SIC – county pair or they can change the industry of an establishment they already operate in the county.

Its also useful to know the characteristics of the firms entering these local retail markets. An established national chain moving into a county is a lot different than a new single unit (mom and pop) store. Jarmin, Klimek and Miranda (2005) distinguish firms by age and geographic scope.

## Measuring the Growth of Young and Small Businesses with the LBD

There has been increased interest in examining the dynamics of small and young businesses. In work supported by the Census Bureau and the Kauffman Foundation, the CES is working to integrate data from the Census Bureau nonemployer business register with the LBD (see Davis et. al., 2005). This work shows that transitions of businesses between the nonemployer and employer business universes are common and important.

**IV.     The Bureau of Labor Statistics Sample Frame:**

**Background**

The Quarterly Census of Employment and Wages (QCEW) is the BLS's sample frame.  The QCEW program is a Federal/State statistical program managed by the BLS and the State Workforce Agencies (SWAs) in the 50 States, the District of Columbia, Puerto Rico, and the Virgin Islands.  The QCEW is used as the sampling frame for most of the BLS' business programs.  Nationwide, in 2004 (third quarter), Federal and State Unemployment Insurance (UI) programs covered 130.2 million full- and part-time workers who received over $1.2 trillion in quarterly  pay.  QCEW data were collected for approximately 8.4 million employer establishments.

The QCEW microdata file contains a wide range of information for each active employer including State UI Account Number (UIN), Establishment Reporting Unit Number (RUN), Federal Employer Identification Number (EIN), six-digit North American Industrial Classification System (NAICS) code, county/township codes, employment for each month during the quarter, total quarterly wages, and the establishment's name(s), addresses, contact name and telephone number.  Three addresses are retained and updated: physical location, U.I. mailing address, and statistical mailing address. Known predecessor and successor relationships are also identified by UI Account Number and establishment Reporting Unit Number (UI/RUN).  The State code, UI Account Number and RUN allow each establishment to be uniquely identified. The unique identification greatly facilitates the accurate linkage of establishments from one quarter to the next.  The purpose of establishment linkage is to identify continuous operations from one quarter to the next, thereby separating them from new businesses and out-of-businesses.

All of the QCEW data originates from employers.  Each State's UI program collects and maintains quarterly administrative record data and business identification information on employers who are subject to State UI laws

All new employers are required to file a "Status Determination Form" to obtain a UI account number and UI tax rate.  Information from this form is also used to assign to each establishment a set of classification codes (industrial, geographical, ownership, and auxiliary).  All covered employers are also required to file a "Quarterly Contribution Report (QCR)".  The employers must report the total wages paid to all workers during the calendar quarter; the wages that are subject to UI taxes; and, the UI taxes that are due to the State.

Also collected in the QCEW, are two types of business names, the legal and trade names. In addition, it retains a Reporting Unit Description (RUD) for each establishment of a multi-establishment employer.  This field further delineates these units by using terminology (store or unit number, plant name, etc.,) provided by the employer, and thus familiar to them, to separately identify each of the establishments in the event that only one of their establishments is selected for a survey.  The QCEW collects type of legal entity (individual, partnership, corporation, and other-limited partnership, S-corporation, household, etc.) that the business or employer represents.

**Data Coverage**

UI coverage is quite broad and comprehensive covering approximately 97 percent of all non-farm wage and salary workers.  All employers who have one or more workers employed on one day in 20 weeks or more in a calendar year or who have a payroll of more than $1,500 in a calendar quarter are required to have their workers covered by UI.  Under the Federal Unemployment Tax Act (FUTA), nonprofit organizations with more than 4 workers during the same time period just cited or meet the same payroll requirement must also cover their workers.  Many States, however, have extended this coverage to the "one or more worker criteria" and thus draw no distinction between profit and nonprofit businesses.

Agricultural coverage also has different coverage requirements from nonagricultural employers.  The agricultural employers are required to be covered if they employ ten or more workers on one day in 20 weeks in a calendar year or have more than a $20,000 payroll in a quarter.  Some of the more densely populated States have also chosen to extend coverage beyond the Federal minimum, resulting in an agricultural coverage rate of 77 percent.  Private households, social clubs, and college fraternities and sororities which employ domestic help and pay wages less than $1,000 in a quarter are excluded from UI laws.  State and local government employers are also required to cover their workers with only a few minor exceptions.

The State UI system identifies and polices new businesses in several other ways.  New businesses report to the UI system.  An average year finds over 1 million new UI numbers assigned.  Nonreporters are found in two ways.  First, cross-matched IRS records may identify a unit with federally reported employment or wages.  Second, the UI claims process self-polices itself.  Each week, over 300,000 new UI claimants seek compensation.  Each of these claims is matched against the UI employer database.  Where there is no match, the UI system seeks out these businesses with substantial powers including freezing bank accounts of businesses that refuse to register.  Each year, over 15 million new claims support the identification of new businesses.

The major exclusions to coverage are the self-employed, workers in selected industries, railroad workers; workers in religious schools, and, selected classes of workers (straight commission life insurance agents, self-employed, unpaid family workers and proprietors).

**Data Collection**

**Status Determination Form (SDF)**

All new employers which become subject to UI coverage are required to file a SDF with the UI unit of the state.  This form is used to determine an employer's tax liability under the State's UI laws and to collect administrative information such as the employer's EIN.  The SDF also requests information for determining industrial activity, county (township in the New England area), and ownership.

After employer liability is determined, the SWA unit assigns a UI account number.  Most new employers are aware of their UI liability and request the SWA to supply a Status Determination Form when they begin their business operations.  Some liable employers, usually small ones, fail to file a SDF.  These employers may be discovered through information on new firms applying

for EINs supplied by the Internal Revenue Service (IRS) to the state each quarter. Other means of discovering liable employers are through the UI claims process; UI field auditor investigations, and an initial filing of the QCR without an UI account number. From sampling and business birth perspectives, this state-specific registration process allows BLS to capture employers whose business operations begin in one state and at a later date, begin operations in other states.

**Quarterly Contribution Report (QCR)**
All liable employers are required to file a QCR with the SESAs for their UI accounts. All of the QCR forms, however, request employment values for each month of the quarter and total wages, taxable wages, and UI taxes due for the quarter. This information and the taxes that are due are necessary for the operation of the UI tax system and also for statistical purposes for the QCEW program. Employers are asked to report, among other items, the total number of covered workers (full and part-time) who earned wages (subject to UI taxes) during the pay period(s) which includes the OMB standard definition of the 12th of each month in the quarter and the total payroll for the quarter. This report is mandatory for employers with a single location as well as employers with multiple locations in the State. The latter group of employers report a summary of these data for all of their establishments covered under the same State UI account on the QCR.

**Multiple Worksite Report (MWR)**
The MWR captures business births and deaths for these multi-establishment employers on an on-going quarterly basis. The MWR is the only source of quarterly data for the individual locations for these multi-establishment employers and thus supports the QCEW for sampling and as an excellent tool for business cycle and policy analysis. Multi-establishment employers with 10 or more employees in the sum of their secondary physical locations and/or industrial activities covered under one UI account, are requested to provide establishment level data using the MWR. Thus, BLS has a universe frame almost entirely at the establishment level. Data collection procedures for multi-establishment employers differ from those for single units. In addition to providing monthly employment and wages. Each quarter, multi-establishment employers are asked to verify the business identifying information (trade name, worksite description, and physical location address) for each establishment (worksite) that is pre-printed on the MWR. New worksites are manually added to the MWR by employers that are then added to the database. Similarly, employers provide the information on deaths or business transfers that are then deleted from that employer's file.

**Report of Federal Employment and Wages (RFEW)**
Federal agencies, whose civilian employees are covered under a separate but comparable UCFE program, do not file QCR forms with State UI programs but instead report employment and wages data directly to the BLS State partner. Since 1993, all States have been using a BLS-designed standardized form, the RFEW that replicates the MWR. Over 50 percent of federal employment is now provided electronically, with all reports to be filed electronically within the next two years.

**The Annual Refiling Survey (ARS)**
The purpose of the ARS is to review and update, if necessary, the classification codes (industrial, geographical, ownership and auxiliary) currently assigned to the establishments stored in the

31

QCEW. The survey is conducted annually with approximately one-third of the establishments. For single establishments, the ARS questionnaire requests that the respondent review an industrial classification statement, or to write in a description of their products and processes. The respondent is also requested to review and update the following, or provide the information if it is not preprinted: 1) physical location address; 2) mailing address; and, 3) county in which the establishment is located. The ARS questionnaire also helps identify new multi-establishment employers by asking if the listed establishment is the only one in that State under that UI number. If no, then the employer provides more information used to determine if the employer should file a MWR form. Through the ARS process, the business and administrative information are reviewed for almost 2 million business establishments each year. Also, potential differences in industry codes, addresses, names, etc that are identified by other BLS programs can be reviewed and reconciled with the information from the QCEW frame.

**Data Processing**

**Current Editing Procedures**
Micro data collected on all of these sources are intensely edited by the State staff and corrected, as necessary. The micro data, including imputed values, are then aggregated to the appropriate QCEW macro-level cells, including size class break-outs for first quarter. The SWA also reviews the macro level data. Both the micro and macro edits include checks for invalid and inconsistent data as well as checks for large and unusual employment and wages fluctuations between and within quarters. The QCEW program uses UI wage records (a file of all individual workers names, social security numbers and wages paid for the quarter) for editing or validating large changes, and determining previously unidentified predecessors and successors. States submit the edited data to the BLS national office approximately four months after the end of the reference quarter.

**Record Linkage Methodology**
The QCEW includes establishment records linked across time. The linkage process used to create and update the database is based upon files that have the same structure across time; these files, therefore, are linked to a new iteration of themselves each quarter. The linkage identifies business establishments which may have gone out-of-business; establishments that remain in business for both periods, and new establishments. Again, state law requires businesses that buy other businesses to report this fact. In addition, a probability-based linkage process is used to identify the small percentage of links for establishments that are missing the appropriate predecessor/successor codes.

**Permanent Random Numbers (PRNs)**
Permanent Random Numbers (PRNs) were developed: 1) to control the sample overlap within periodic surveys from one period to the next; 2) to minimize the sample overlap, to the extent possible, between surveys at any given time; and 3) to spread the burden imposed on small businesses by reducing the probability that they will be included in more than one Bureau survey. The spreading of burden is accomplished by assigning each major survey a start point within a stratum, and then having the survey select sample units sequentially from that point. Randomness is retained as the PRNs were randomly assigned to the QCEW files beginning in 1995 and

subsequently to all new establishments on each quarterly file.  The PRNs for each new establishment within a quarter are collocated within defined strata.  The collocation process spreads the PRNs evenly and hence the new businesses within each stratum.  Thus, all the surveys have proper representation of continuous establishments and new businesses in the sample and thereby reducing a source of potential bias that can arise if new establishments are different from continuous ones.  The establishments that have gone out-of-business are removed along with their PRN's.  The PRN's are also assigned at the UI account level to accommodate the CES Program for which the sampling unit is U.I., for the other programs sampling unit is the establishment.

**Data Outputs and Uses**

**General**
The QCEW is the most complete and timely source of monthly employment and quarterly wages information by detailed industry and geography.  Consequently, QCEW data are used extensively in many economic and statistical applications.  These include UI program administration, macro-economic research, survey employment benchmarking.  The Bureau of Economic Analysis uses the QCEW wages for about 60% of the quarterly Personal Income component of the Gross Domestic Product at national, state and county levels.  BLS recently completed a four-week acceleration in the data review and submittal process to support BEAs improvements to the Personal Income series..

**Sampling Frame for BLS Surveys**
The QCEW is the sampling frame for the following surveys: Current Employment Statistics (CES), Occupational Employment Statistics (OES), Producer Price Index (portions), Occupational Safety and Health, National Compensation Survey, Productivity programs, the Job Openings and Labor Turnover survey.  It is also the employment benchmark for the monthly CES payroll survey, OES, JOLTS and other programs.

The QCEW receives new business births very promptly. Each quarter, BLS obtains extracts of the current and preceding quarters' data from the state UI files for use in the QCEW program. For the most recent data available, 86 percent of the new business births that are included in the BLS QCEW opened within the same quarter that they were liable for UI coverage.

The CES program draws an updated sample every year supplemented by a sample of new businesses six months later.  The OES program draws a new sample on a semi-annual basis and the JOLTS program draws an updated sample every year.  These updated and new samples better represent the target population through updates to the sampling frame for new businesses, out-of-businesses, and realignment of continuous businesses within each sampling stratum.  The sample overlap is controlled through permanent random numbers techniques.  An update to the sample also includes updating of addresses for old sampling units thereby reducing non-respondents.

**Business Employment Dynamics Statistics**
The QCEW is also used to produce the new Business Employment Dynamics data series to analyze the job creation and job destruction process by industry, size, age and geography.  The consistent detail available quarterly allows detailed data for business cycle analysis, among other

studies. BLS recently completed the development that can be used to analyze the job creation and job destruction process. Data from the first quarter of 1990 forward are now available to researchers for qualified projects.

**Use by the Census Bureau**

The QCEW is used by the U.S. Census Bureau in two ways. The QCEW micro-level data is the basis for the Census Longitudinal Employer Household Dynamics (LEHD) program. The detailed establishment level reporting provides the accuracy needed to properly distribute establishments, employment and wages to the county level. The QCEW EIN field provides the key matching variable to other data sources available to Census. Also, BLS shares over 1 million NAICS codes and physical location addresses with the Census Bureau each year minimizing employer burden, reducing Census costs for mailing, handling, coding and postage and, importantly, improving comparability among the two agencies data outputs. Most of these codes are for new and small businesses.

**State and Local Area Usage**

At the state level, the QCEW is used for analysis, economic development, policy analysis, transportation planning, and a wide range of other purposes, including as a sample frame for state local areas surveys such as Job Vacancy Surveys.

**Summary**

The QCEW program has many strengths. It is quarterly, thereby achieving a smooth, consistent series of measures. Monthly employment (collected quarterly) allows instant determination of size for sampling and understanding of seasonally patterns for weighting. Monthly employment also allows close quarterly tracking of the monthly payroll employment survey versus the universe and detailed studies of differences by several characteristics.

Coverage includes both public and private sectors for virtually universal coverage. Only the self employed and a few smaller industrial sectors are missing. The self employed could be addressed through data sharing.

The State UI system requires that businesses report to each state in which they have employees. This feature insures that the geographic allocation of employment, wages and samples are very accurate to the state level. Further allocations to MSA and county levels (or other levels) are supported by the alignment to the state level and the quarterly MWR collection.

**V.      Summary**

This paper discusses the treatment of births and deaths in the Quarterly Census of Employment and Wages (QCEW) and Current Employment Survey (CES) program of the BLS and in the Census Bureau's Business Register (BR); the Retail, Wholesale, and Services surveys; and the Longitudinal Business Database (LBD). Each has a different primary purpose, its own way to identify births and deaths, and its unique methodology for treating them.

**Primary Purposes**

The primary purpose of the BLS CES program is to estimate monthly change in the level of U.S. nonfarm employment, hours, and earnings for detailed industry and geographic levels. The BR provides an up-to-date universe of U.S. businesses, including basic information for each business. The monthly Retail and Wholesale surveys estimate percent change in total U.S. sales/receipts and inventories for detailed industry levels while the annual Retail, Wholesale, and Services surveys provide annual level estimates of sales/receipts, inventories, expenses, and other items for detailed industries. The Census Bureau's LBD provides a research data set that can be used to study business dynamics.

**Identifying Births and Deaths**

The BLS CES program births and deaths are identified in terms of State Unemployment Insurance (UI) tax accounts opened or closed since the most recent sampling frame was set. The BR identifies them as new and dead establishments, new and inactive Employer Identification Numbers (EINs), and Social Security Numbers (SSNs). Births and deaths for the Retail, Wholesale, and Services surveys are identified in terms of new and inactive noncertainty EINs and organizational changes to certainty company units that result in adding or subtracting establishments from the company units. The Census Bureau's LBD allows users to tailor the way births and deaths are identified to meet their particular research needs.

**Treatment of Births and Deaths**

The CES program uses a new sample every year. The program produces monthly estimates of change in the level of employment, hours, and earnings from the UIs included in the sample. Data are imputed for all nonrespondents, including deaths. Since the imputation procedure does not totally account for employment changes due to births and deaths, the estimates are supplemented with estimates of net birth/death residual employment. The supplemental forecasts are obtained from ARIMA model projections based on 5 years of historical net birth/death residuals.

Birth and death updates for the BR come from various Census Bureau survey and census data collections and other government administrative data sources.

The Retail, Wholesale, and Services surveys select new samples about once every five years. Administrative payroll tax filing requirement information and other data are used to determine new EIN births and deaths. Once each quarter EIN births are sampled and EIN deaths are identified. These are incorporated into the ongoing samples on a monthly basis. To account for lags in introducing births into the monthly surveys, sales are imputed for noncertatinty EINs that are known to be inactive though administrative records indicate otherwise. For the annual surveys, the Census Bureau collects data from EIN births selected in the first three quarterly operations of the year after the survey reference year. Changes for certainty EINs and company units are identified and applied on a continual basis.

The Census Bureau's LBD is based off of the BR and allows researchers the freedom to identify births and deaths in a manner consistent with their particular research purposes.

**Common Concerns**

While there are differences in the primary purposes of the programs discussed in this paper and in the ways in which each defines and treats births and deaths, BLS and Census have several common concerns. Among them are these:

a. Timeliness of the input data
b. Quality of the input data
c. Ability to distinguish true births from changes to existing units
d. Adequacy of industry classification
e. Adequacy of methods used to account for births and deaths in business surveys

## VI.      References

Bartelsman, E., Scarpetta, S., Schivardi, F. (2003), "Comparative Analysis of Firm Demographics and Survival: Micro-Level Evidence for The OECD Countries", OECD Economics Department, Working Paper No. 348.

Black, J. (2001), "Changes in Sampling Units in Surveys of Businesses," Federal Committee Statistical Methodology (FCSM) Research Conference: 2001, FCSM Statistical Working Policy Paper 34.

Burton, J.N., King, C.S., Hunt, J.W. (2001), "Benefits and Limitations of Using Only Administrative Data to Update Current Business Surveys with New Employer Births," Joint Statistical Meetings of the American Statistical Association, August 2001.

Davis, S., Haltiwanger, J., Schuh, S. (1996), <u>Job Creation and Destruction</u>, Cambridge: MIT Press.

Davis, S., Haltiwanger, J., Jarmin, R., Krizan, C.J, Miranda, J., Nucci, A., Sandusky. L.K. (2005), "Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes", presented at the "NBER/CRIW Conference on "Producer Dynamics: New Evidence from Micro Data", April 8-9, 2005.

Dunne, T., Roberts, M., Samuelson, L. (1988), "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries", RAND Journal of Economics, Vol. 19, No. 4, pp. 495-515.

"Employment and Earnings Technical Notes," (February 2005), Bureau of Labor Statistics.

Getz, P., Kropf, J., Strifas, S. (1997), "Measuring the Contribution of Business Births and Deaths to Overall Employment Movements," Proceedings of the American Statistical Association, Summer 1997.

Gillott, L.M., Burton, J.M., King, C.S. ((2003), "The Use of Administrative Data to Compute Measure of Size for New Employer Births," Federal Committee Statistical Methodology (FCSM) Research Conference: 2003, FCSM Statistical Working Policy Paper 37, pp. 43-48, www.fcsm.gov

Jarmin, R. Klimek, S., Miranda, J. (2005), "The evolution of retail markets in micropolitan, metropolitan and rural regions", presented at the "NBER/CRIW Conference on "Producer Dynamics: New Evidence from Micro Data", April 8-9, 2005.

Jarmin, R., Miranda, J.,"The Longitudinal Business Database", Center for Economic Studies, Working Paper No. 02-17

King, C.S. and Struble, R.E. (2001), "Procedures to Account for Entries in Business Surveys," Proceedings of the 2001 Methodology Symposium, www.statcan.ca

Kropf, J., Strifas, S., Traetow, M. (2002), "Accounting for Business Births and Deaths in CES: Bias versus Net Birth/Death Modeling," Proceedings of the American Statistical Association, Summer 2002.

Stamas, G., Goldenberg, K., Levin, K., Cantor, D. (1997), "Sampling for Employment at New Establishments in a Monthly Business Survey," Proceedings of the American Statistical Association, Summer 1997.

U.S. Census Bureau (2004), Current Business Reports, Series BR/03-A, Annual Benchmark Report for Retail and Food Services: January 1992 Through February 2004, Washington, D.C.

Walker, E. (1997), "The Census Bureau's Business Register: Basic Features and Quality Issues," American Statistical Association Meetings, August 1997.

**Questions for FESAC Committee**

1. Are there possible model enhancements that FESAC would recommend that BLS examine? Additional variables or different modeling techniques?

2. Are there methods other than modeling that FESAC would recommend BLS examine to account for business birth/death employment in the CES program?

3. Are there additional evaluation techniques recommended for measuring the performance of the BLS model-based approach to birth/death employment contributions?

4. Are there ways to measure the adequacy of methods/assumptions used to account for births and deaths? Are the measures taken by CES sufficient?

5. Adding births is important for ensuring that samples are representative of the intended universes. However, the data for births added each month to the retail and wholesale surveys do not make much difference in the estimate of month-to-month percent change in sales for total retail and wholesale. Census is thinking of researching the implications of doing the following:
   - continue the quarterly update process to the samples
   - do not collect data from births added to the monthly surveys until they have first been mailed an annual survey questionnaire.
   - use administrative data to impute for the births in the monthly surveys during the interim.

This approach would reduce mailing costs and burden slightly and would eliminate some editing challenges. Since the first actual data collection for the births would be done via a mandatory survey, it is believed that their response to the monthly surveys may be increased.

Does this seem like a reasonable approach? What issues should be considered in such a study? Is there another approach that you would recommend?

6. The Census Bureau has used its quarterly birth sample to improve the industry codes for the births subjected to sampling. Are there alternative sources for industry codes that may be less costly, of equal or higher quality, and no less timely?

7. What steps should the agencies take to improve the accuracy and usefulness of longitudinal files, such as the LBD, for tracking producer dynamics?

| Form **SS-4**<br>(Rev. December 2001)<br>Department of the Treasury<br>Internal Revenue Service | **Application for Employer Identification Number**<br>(For use by employers, corporations, partnerships, trusts, estates, churches,<br>government agencies, Indian tribal entities, certain individuals, and others.)<br>▶ See separate instructions for each line.  ▶ Keep a copy for your records. | EIN _____<br><br>OMB No. 1545-0003 |

**Type or print clearly.**

**1** Legal name of entity (or individual) for whom the EIN is being requested

**2** Trade name of business (if different from name on line 1)  |  **3** Executor, trustee, "care of" name

**4a** Mailing address (room, apt., suite no. and street, or P.O. box)  |  **5a** Street address (if different) (Do not enter a P.O. box.)

**4b** City, state, and ZIP code  |  **5b** City, state, and ZIP code

**6** County and state where principal business is located

**7a** Name of principal officer, general partner, grantor, owner, or trustor  |  **7b** SSN, ITIN, or EIN

**8a** Type of entity (check only one box)
- ☐ Sole proprietor (SSN) _____
- ☐ Partnership
- ☐ Corporation (enter form number to be filed) ▶ _____
- ☐ Personal service corp.
- ☐ Church or church-controlled organization
- ☐ Other nonprofit organization (specify) ▶ _____
- ☐ Other (specify) ▶

- ☐ Estate (SSN of decedent) _____
- ☐ Plan administrator (SSN) _____
- ☐ Trust (SSN of grantor) _____
- ☐ National Guard  ☐ State/local government
- ☐ Farmers' cooperative  ☐ Federal government/military
- ☐ REMIC  ☐ Indian tribal governments/enterprises
- Group Exemption Number (GEN) ▶ _____

**8b** If a corporation, name the state or foreign country (if applicable) where incorporated  |  State  |  Foreign country

**9** Reason for applying (check only one box)
- ☐ Started new business (specify type) ▶ _____
- ☐ Hired employees (Check the box and see line 12.)
- ☐ Compliance with IRS withholding regulations
- ☐ Other (specify) ▶

- ☐ Banking purpose (specify purpose) ▶ _____
- ☐ Changed type of organization (specify new type) ▶ _____
- ☐ Purchased going business
- ☐ Created a trust (specify type) ▶ _____
- ☐ Created a pension plan (specify type) ▶ _____

**10** Date business started or acquired (month, day, year)  |  **11** Closing month of accounting year

**12** First date wages or annuities were paid or will be paid (month, day, year). **Note:** If applicant is a withholding agent, enter date income will first be paid to nonresident alien. (month, day, year) . . . . . . . . . . . . . . . ▶

**13** Highest number of employees expected in the next 12 months. **Note:** If the applicant does not expect to have any employees during the period, enter "-0-." . . . . . . . . . . . ▶

| Agricultural | Household | Other |
|---|---|---|
| | | |

**14** Check **one** box that best describes the principal activity of your business.
- ☐ Construction  ☐ Rental & leasing  ☐ Transportation & warehousing  ☐ Health care & social assistance  ☐ Wholesale–agent/broker
- ☐ Real estate  ☐ Manufacturing  ☐ Finance & insurance  ☐ Accommodation & food service  ☐ Wholesale–other  ☐ Retail
- ☐ Other (specify)

**15** Indicate principal line of merchandise sold; specific construction work done; products produced; or services provided.

**16a** Has the applicant ever applied for an employer identification number for this or any other business? . . . . ☐ Yes  ☐ No
**Note:** If "Yes," please complete lines 16b and 16c.

**16b** If you checked "Yes" on line 16a, give applicant's legal name and trade name shown on prior application if different from line 1 or 2 above.
Legal name ▶ _____  Trade name ▶ _____

**16c** Approximate date when, and city and state where, the application was filed. Enter previous employer identification number if known.

| Approximate date when filed (mo., day, year) | City and state where filed | Previous EIN |
|---|---|---|
| | | |

| | Complete this section **only** if you want to authorize the named individual to receive the entity's EIN and answer questions about the completion of this form. | |
|---|---|---|
| **Third Party Designee** | Designee's name | Designee's telephone number (include area code)<br>(    ) |
| | Address and ZIP code | Designee's fax number (include area code)<br>(    ) |

Under penalties of perjury, I declare that I have examined this application, and to the best of my knowledge and belief, it is true, correct, and complete. | Applicant's telephone number (include area code)<br>(    )

Name and title (type or print clearly) ▶

Signature ▶ _____  Date ▶ _____ | Applicant's fax number (include area code)<br>(    )

**For Privacy Act and Paperwork Reduction Act Notice, see separate instructions.**  Cat. No. 16055N  Form **SS-4** (Rev. 12-2001)