Impact of business births and deaths in the payroll survey

The CES probability-based sample redesign accounts for most business birth employment through the imputation of business deaths, with the remaining portion estimated by a net birth/death model

Kirk Mueller

The Current Employment Statistics (CES) survey, conducted by the Bureau of Labor Statistics (BLS), is a monthly survey of more than 400,000 business establishments. The CES program provides estimates on employment, hours, and earnings by industry detail for the Nation, States, and metropolitan areas. The CES is widely considered one of the most timely and accurate economic indicators published by the Federal Government.

The CES sample-based employment estimates for March of each year are benchmarked, or reanchored, annually to the March universe count derived principally from the Quarterly Census of Employment and Wages (QCEW) program. These QCEW population counts are much less timely than sample-based estimates and are used to provide an annual point-in-time census for employment. For national series, only the March sample-based estimates are replaced with the population counts.[1]

BLS completed a comprehensive redesign of the CES sample in 2003, changing the survey from a quota-based sample to a probability-based sample.[2] The probability-based sample redesign addressed one of the major limitations of the previous quota-based sample: the absence of a method to directly measure new business births. The new probability-based sample accounts for most business birth employment through the imputation of business deaths, with the remaining portion estimated by a net birth/death model that calculates the effect of the imputation, measures the imputation error, and generates a forecast of this error to adjust the current estimate.

With the introduction of the redesign, many questions have arisen with respect to the new model-based estimation of the net of births and deaths. This article discusses the underlying assumptions of the model and the rationale behind them. It also discusses the reasons why the total accounting for business births is cyclically sensitive, in spite of the use of the forecast net birth/death values. Lastly, it draws comparisons between the bias adjustment model of the quota sample design and the net birth/death model of the probability sample design. The models differ in the portion of the population that they are meant to measure.

Probability sample design

From its inception in the 1930s until the redesign, the payroll survey was collected as a quota-based sample. A bias adjustment model was used to account for the employment movement each month not captured by the sample, including employment growth because of the birth of new establishments. Over time, both internal and external reviews of the CES program concluded that a probability-based sample would benefit the program by introducing a more standard survey design and decreasing the reliance of model-based adjustments. After several years of research, BLS began in June 2000 to implement in the CES a probability-based sample, phased in by industry. This process was

Kirk Mueller is a supervisory statistician in the Division of Current Employment Statistics, Bureau of Labor Statistics. E-mail: mueller.kirk@bls.gov

completed with the 2002 benchmark release in June 2003, when the last industry was converted to a probability basis.

During the redesign phase in period, BLS conducted production tests of the new sample design and made parallel estimates each year for industries prior to their official implementation with the subsequent benchmark. The new survey design includes both the new sample composition and the use of a two-step process to account for the employment associated with business births. This process first imputes a portion of the birth employment from the employment associated with business deaths. The second step models the historical difference between the imputation and the actual relationship between business birth and business death employments; this step is referred to as the net birth/death model.

The establishments that make up the population of interest for the CES can be broken up into three segments relative to a benchmark month. Those segments are 1) establishments that continue to employ workers after the benchmark, 2) establishments that go out-of-business after the benchmark, and 3) establishments that begin to employ workers after the benchmark. With the probability sample implementation, the entire population of establishments that are in business in the benchmark month is appropriately represented for the period from which the sample is selected. As a result, the sample accounts for the first two segments—establishments that are continuing in nature and establishments that go out of business. The population employment for these units is moved forward through the use of weighted link relative estimation. If it is assumed that there is little difference in the response rates between the continuing units in the sample and the units that go out of business in the sample, then the monthly estimate based on the establishments in the sample should appropriately capture the employment movements within these two segments of the universe of establishments. (The accuracy of the assumption of similar response rates is addressed later in the article.) Accounting for the third segment—establishments that are births—still requires additional steps.

To understand how the employment estimates are moved forward with the weighted link relative, it is necessary to consider the form of the estimator. The basic formula for estimating all employees is:

$$\hat{AE}_c = \left(\hat{AE}_p \times \frac{\sum_i \left(w_i \times ae_{c,i} \right)}{\sum_i \left(w_i \times ae_{p,i} \right)} \right)$$

where

$\hat{AE}_c$ = estimate of all employees for the current month
$\hat{AE}_p$ = estimate of all employees for the previous month
i = the *i-th* sample unit
w_i = the weight for the *i-th* sample unit
$ae_{c,i}$ = *i-th* sample unit that reports for the current month
$ae_{p,i}$ = *i-th* sample unit that reports for the previous month

The estimator requires that the business reports data for both the previous and current months if its data are to be used. This is referred to as a matched sample. This estimator uses the trend in the matched sample to move the previous month's estimated employment forward.

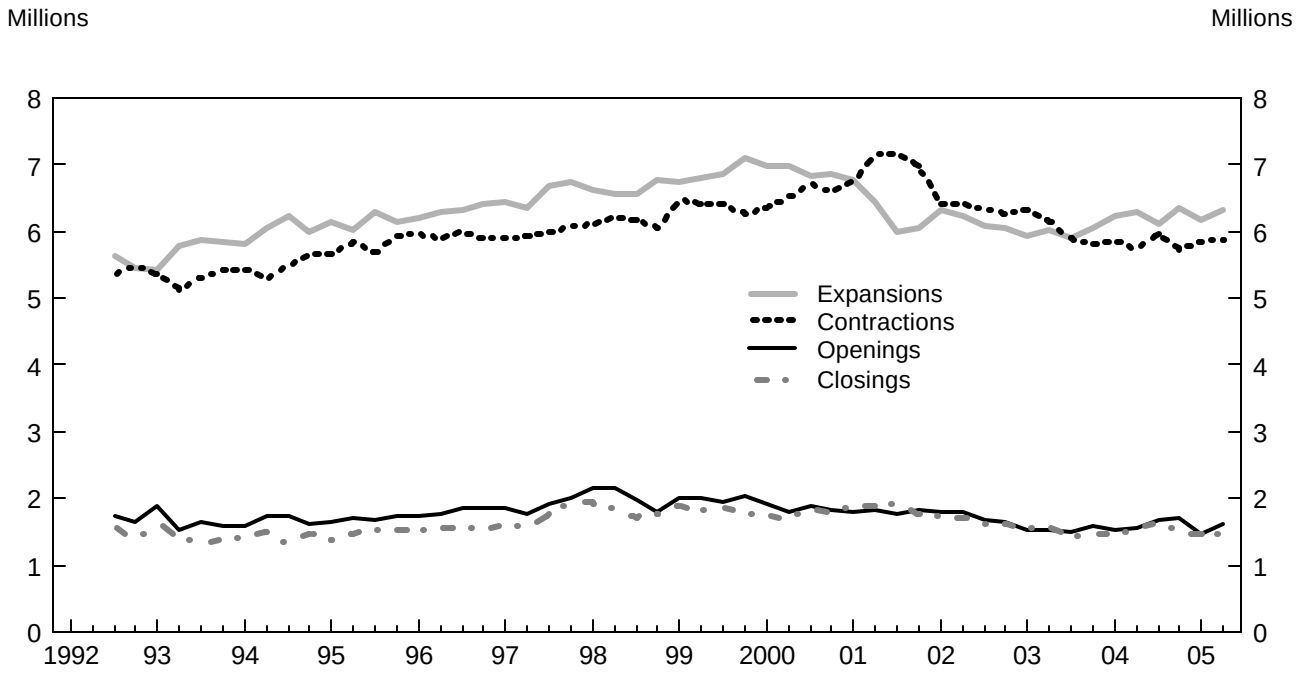
As mentioned previously, the third segment of the CES population of interest is establishments that open after the benchmark, or business births. Like many establishment surveys, the CES has difficulty with obtaining and developing a timely sample frame for business births. This is because a lag exists between an establishment opening for business and its appearance on the universe frame, where it would be available for sampling. This lag currently is about 7 months. In contrast, the lag from the reference month to the CES first publication of the employment estimates for that reference month is 3 weeks. Absent a sample, some form of modeling of this employment is necessary to account for the business births and their associated employment growth.

Early research associated with the redesign efforts indicated a strong relationship between the employment associated with business births and that associated with business deaths. This relationship can be seen in the Business Employment Dynamics (BED) data when a comparison is made between the employment associated with quarterly business openings and that associated with quarterly business closings.[3] (See chart 1.) The primary difference between the BED data and the CES concepts is that the BED data track births over a quarter. The birth employment relevant to the CES is the employment associated with births since the last benchmark.

Imputation of deaths

As was mentioned earlier, the probability sample with the weighted link relative estimator will accurately represent the movement of the population. However, as a practical matter, units that have gone out of business generally do not report their data for the month in which they go out of business. There are two primary reasons. First, the CES is a voluntary survey and not all firms respond. Second, companies by definition have no employees once they are out of business, and there may be no one to report their data.[4] As a result, these unreported sample events are not included in the calculation of the weighted link relative and are counter to the assumption discussed earlier that there were no large differences between the response rates of continuous units and those that go out of business.

Chart 1. **Private sector components of gross job gains and gross job losses, Business Employment Dynamics data, September 1992–June 2005, seasonally adjusted**



To resolve this, a rather intensive followup with all nonresponders would be required each month to attempt to distinguish between nonrespondents and business deaths. Even if this could be accomplished successfully, a model would still be needed for all of the employment associated with business births.

Instead, BLS decided to use the underlying relationship between business birth employment and business death employment that was described in the previous section to at least partially account for the employment associated with business births. Rather than identifying all the deaths for the estimation process, the logic is adjusted to exclude all business deaths from the sample link. Deaths that are nonrespondents are automatically excluded from the matched sample as they have no current month data, and establishments that report that they are out of business are treated as nonrespondents for the current month and are also excluded from the link. As a result, the link calculation is based solely on continuing units. While this first step accounts for a large portion of the birth employment, it does not account for it all.

To understand what is occurring with this first step, it is helpful to break down the estimation process. Conceptually, the previous month's employment can be broken into two parts: 1) the previous month's employment for firms that continued to employ workers in the current month, and 2) the previous month's employment for firms that go out of business, that is report zero employment, in the current month. Next, consider the application of the weighted current-to-previous month employment ratio for sample units to the previous month employment level for each of the two pieces. This weighted ratio is the same used in the weighted link relative estimator. The employment at establishments that employed workers the prior month and continue to employ workers during the current month is moved forward by the link of the continuous sample units. Applying the link of the continuous units to the employment associated with business deaths effectively imputes a level of birth employment and the growth of previous births for the current month.

Based on the relationship shown in the BED data, this imputation should account for a large portion of the business birth employment. The degree to which this imputation is different from the birth employment is left to be modeled and represents the second step in the accounting for business birth employment.

Modeling residual birth employment

The business birth employment not accounted for by the imputation of business deaths in the sample is modeled as an AutoRegressive Intergrated Moving Average (ARIMA) time

series.[5] This model is referred to as the net birth/death model. The Bureau's Longitudinal Database (LDB) is the basis for developing the historical relationship to be modeled. The LDB links establishments over time, which allows for the identification of the continuous establishments, establishments that go out of business, and births of establishments. To develop the history for modeling, the same handling of business deaths as described for the CES sample data is applied to the population data. Establishments that go out of business have employment imputed for them based on the rate of change of the continuous units. The employment associated with continuous units and the employment imputed from deaths are summed. The difference is compared with the actual population level to create the series modeled by the birth/death models.[6]

To date, the residual net birth/death component has shown to be a relatively stable portion of the population employment regardless of the point in the business cycle. This may seem counterintuitive until the impact of the imputation of the business deaths is considered in more detail. However, the BED data show that the majority of the employment change in the population is explained by changes in the continuing units rather than the relationship between the employment associated with business births and that associated with business deaths. Furthermore, the BED birth/death relationship is somewhat cyclical like the continuous unit population. The CES application of this relationship takes a step further with the application of the continuous unit link to the employment associated with business deaths. This imputation of employment from business deaths does not provide an exact one-to-one relationship between employment from establishment births and deaths; rather, it is dependant upon the movement of the continuous units.

To complete the estimation formula for the entire population, a net birth/death value must be added to the weighted link relative described earlier. The final formula is as follows:

$$\hat{AE}_c = \left(\hat{AE}_p \times \frac{\sum_i \left(w_i \times ae_{c,i} \right)}{\sum_i \left(w_i \times ae_{p,i} \right)} \right) + \text{net birth/death}$$

Cyclical sensitivity

The total business birth employment is accounted for by both the imputation of business deaths and the net birth/death value. While the net birth/death value is a fixed, projected value, the imputation of deaths is dependent upon current sample information.

The impact of the imputation is best seen in the following series of examples. (See table 1.) For the examples, assume that there is full response associated with the sample. This allows for the evaluation of the imputation process, without the complication of nonresponse. In these examples, the sample is broken down to illustrate the impact of the imputation of business deaths and the estimate of total business birth employment. For each of these examples, assume that the previous month's employment level is 20,000 and the net birth/death factor for the cell is 0.

In each of the examples, a table shows the previous month reported employment for each sample reporter in column 1, the current month reported employment for each sample reporter in column 2, the sample weight for each sample reporter in column 3, the previous month weighted employment (weight times the previous month employment) for each sample reporter in column 4, and the current month weighted employment (weight times the current month employment) for each sample reporter in column 5. The last two rows in each table show the weighted previous and current month employment totals for the sample both including the sample member that went out-of-business and excluding the sample member.

For each example, three calculations are computed.

1) An estimate **including** the sample member that goes out of business (report zero employment)
2) An estimate **excluding** the sample member that goes out of business (this estimate is the CES estimate based upon the existing estimation algorithm)
3) A calculation of the employment associated with the imputation of business deaths (it is the difference between the two estimates listed above)

The differences in the total birth employments in the three examples illustrate the impact of the imputation of business deaths. They also indicate that the total accounting for business birth employment is sensitive to current business cycle information. The first example provides a case where the continuous units are relatively flat. (See table 1, example a.) Applying the link relative estimation formula with the sample death included results in an employment estimate of (20,000*7,590.10/7,840.90) = 19,360. Applying the link relative estimation formula with the sample death excluded results in an employment estimate of (20,000*7,590.10/7,502.00) = 20,235. In this case, the imputation of the death added 875 in employment associated with business births.

In the second example, the continuous units are expanding. (See table 1, example b.) Applying the weighted link relative estimation formula with the sample death included results in an employment estimate of (20,000*8,425.40/7,840.90) = 21,491. Applying the weighted link relative estimation formula with the sample death excluded results in

Table 1. Example a: continuous units are relatively flat

Previous month sample employment	Current month sample employment	Sample weight	Previous month weighted employment	Current month weighted employment
100	103	7.20	720.00	741.60
70	71	11.12	778.40	789.52
400	411	6.80	2,720.00	2,794.80
8	9	68.78	550.24	619.02
4	7	1.00	4.00	7.00
3	2	11.12	33.36	22.24
11	8	6.80	74.80	54.40
35	38	32.00	1,120.00	1,216.00
70	63	11.12	778.40	700.56
65	58	11.12	722.80	644.96
5	0	67.78	338.90	.00
...	...	...	[1]7,840.90	[1]7,590.10
...	...	...	[2]7,502.00	[2]7,590.10
Example b: continuous units are growing				
100	108	7.20	720.00	777.60
70	76	11.12	778.40	845.12
400	416	6.80	2,720.00	2,828.80
8	14	68.78	550.24	962.92
4	12	1.00	4.00	12.00
3	7	11.12	33.36	77.84
11	13	6.80	74.80	88.40
35	43	32.00	1,120.00	1,376.00
70	68	11.12	778.40	756.16
65	63	11.12	722.80	700.56
5	0	67.78	338.90	.00
...	...	...	[1]7,840.90	[1]8,425.40
...	...	...	[2]7,502.00	[2]8,425.40
Example c: continuous units are declining				
100	98	7.20	720.00	705.60
70	65	11.12	778.40	722.80
400	405	6.80	2,720.00	2,754.00
8	4	68.78	550.24	275.12
4	2	1.00	4.00	2.00
3	1	11.12	33.36	11.12
11	8	6.80	74.80	54.40
35	37	32.00	1,120.00	1,184.00
70	58	11.12	778.40	644.96
65	63	11.12	722.80	700.56
5	0	67.78	338.90	.00
...	...	...	[1]7,840.90	[1]7,054.56
...	...	...	[2]7,502.00	[2]7,054.56

[1] Total with business death included.

[2] Total without business death included.

an employment estimate of (20,000*8,425.40/7,502.00) = 22,462. In this case, the imputation of the death added 971 in business birth employment.

In the last example, the continuous units are contracting. (See table 1, example c.) Applying the link relative estimation formula with the sample death included results in an employment estimate of (20,000*7,054.56/7,840.9) = 17,994. Applying the link relative estimation formula with the sample death excluded results in an employment estimate of (20,000*7,054.56/7,502.00) = 18,807. In this case, the imputation of the death added 813 in employment.

In other words, the employment level from the imputation differs depending upon the movement of the continuous sample. This is important when considering how the CES ac-

counts for business births and their employment. While the net birth/death figure is a forecasted value, there is current information being used through the imputation of business deaths. As a result, there is sensitivity to current economic conditions in the assumptions for accounting for business birth employment.

If each of these examples is changed to assume that there are sample nonrespondents, then the total amount of birth employment accounted for by each example's imputation is a lower bound for the total births employment in the population. This is because some of the nonrespondents may be deaths with employment imputed for them. For the nonrespondents that are still in business, the link from continuous units is appropriate. If some of the nonrespondents are out of business, then in the estimation process, they also have employment imputed for them.

Bias adjustment vs. net birth/death

Historically, the CES has relied on modeling of some segment of the population to complete the most accurate and current employment picture possible. Under the old quota-based design, which was discontinued in 2003, this modeling was referred to as bias adjustment. A comparison between the bias adjustment and the birth/death adjustment is frequently made by CES data users. However, there are several distinctions between the two models. Both models account for the only nonsample-based adjustment to the CES estimates; however, the birth/death model is not simply an improved bias adjustment model. Bias adjustment was a total error correction model that was used to account for several deficiencies in the quota sample including a nonrandom sample and response errors. As a result, the bias adjustment models were directly driven by revisions to the estimates with the previous benchmark and assumed all error and variability in the estimate should be corrected by the model.

Under the current probability-based design of the CES survey, only the business births are not directly accounted for through the sample design. The residual net birth/death model can have error associated with it that is not directly tied to benchmark revisions. The model values are affected by defined portions of the population—business births and business deaths. Benchmark revisions can be attributable to nonresponse error, reporting error, sample error or simple sample variability, and the error associated with the modeling for the net of births and deaths. With the new design, each of these components can be examined separately and corrected as the need arises. As a result, it is possible for net birth/death factors to increase in industries with downward benchmark revisions or in industries with upward revisions.

The bias model and the birth/death model are expected to capture different portions of the population movement and, under the current survey, more of the population movement is captured over time through the sample and less is captured through modeling. With the introduction of the new design, parallel estimates were made for a 12-month period in each division. Official quota-based estimates and the probability-based estimates performed similarly; however, generally less birth/death adjustment was applied to the probability estimates than was applied by the bias adjustment model used with the quota-based sample. (See table 2.)

Birth/death model performance

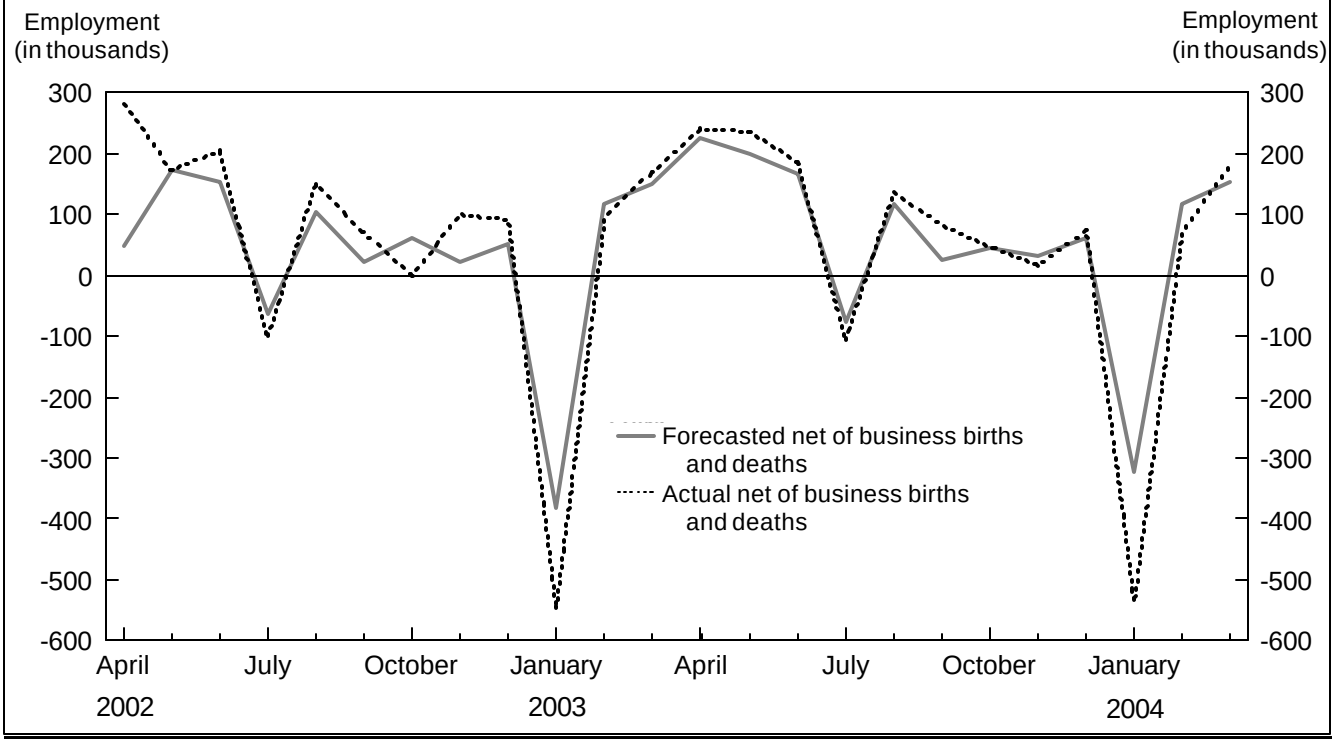
With the full conversion of the CES sample to a probability and NAICS basis, an analysis of the performance of the birth/death model against population data can be performed with the refitting of the models each year. While benchmark revisions have been small in recent years, it is possible that the small revisions could be a result of offsetting errors within the CES-estimation process rather than the quality of the birth/death model. An examination of the forecasted net birth/death factors compared with the actual net of business births and deaths shows that the two did not differ greatly for the April 2002–March 2004 period at the total private level. (See chart 2.)

Table 2. Bias and birth/death factors for parallel estimation periods

Industry	Industry implemented	12-month bias total	12-month birth/death total
Mining	June 2001	0	–8,000
Construction	June 2001	144,000	119,000
Manufacturing	June 2001	96,000	17,000
Wholesale trade	June 2000	153,000	37,000
Retail trade	June 2002	221,000	87,000
Transportation and public utilities	June 2002	90,000	23,000
Finance, insurance, and real estate	June 2002	25,000	9,000

NOTE: Services estimates were not produced in parallel because of the conversion to NAICS in June 2003.

Chart 2. **Total private forecasted versus actual net of business births and deaths, not seasonally adjusted, April 2002–March 2004**



Additional research

The complete accounting of business birth employment does contain a cyclical component that results from the imputation process, and analysis of both the benchmark revisions and the comparisons of the birth/death factors with population data indicates that the imputation and model combination are performing well. However, there are no variables in the net birth/death model that provide information that is more current than the most recent benchmark. Future research with respect to the birth/death model will involve the examination of variables that can be incorporated on a concurrent or lagged basis. These variables may provide more recent information than what is currently present in the model.

Notes

[1] The Bureau's unemployment insurance (UI) universe count is a quarterly tabulation, from administrative records, of the number of employees covered by unemployment insurance laws. UI universe counts, available on a lagged basis, contain individual employer records for more than 8 million establishments and cover a little more than 97 percent of total nonfarm employment; they thus provide a benchmark for the sample-based estimates. For the small segment of the population not covered by UI, BLS develops employment benchmarks from several alternative sources. More information on benchmarking of the CES estimates can be found on the Internet at **http://www.bls.gov/web/cesbmart.htm**.

[2] A probability-based sample is selected through a random process, and the probabilities of selection are known for each unit in the population. A quota-based sample is derived through a sampling process that is repeated, until a minimum responding sample, or quota, is obtained for each characteristic of interest. Details on the implementation of the CES redesign are available in an article by Sharon Strifas, "Revisions to the Current Employment Statistics National Estimates Effective May 2003," *Employment and Earnings*, June 2003, pp. 3–19.

[3] The Business Employment Dynamics data are a set of statistics generated from the Quarterly Census of Employment and Wages, or ES–202, program. These quarterly data series consist of gross job gains and gross job losses statistics from 1992 forward. These data help to provide a picture of the dynamic state of the labor market. More information on the Business Employment Dynamics data can be found on the Internet at **http://www.bls.gov/bdm/home.htm**.

[4] Exceptions occur when all worksites are reported in either an aggregate single report or to the Electronic Data Interchange Center; then the location going out-of-business is reported.

[5] ARIMA modeling uses lags and shifts in the historical data to uncover patterns, such as moving averages and seasonality.

[6] More detailed technical model descriptions have been published in the Statistical Proceedings of the American Statistical Association and are available on the Internet at **http://www.bls.gov/ore/abstract/st/st020090.htm**.