MEASURING THE CONTRIBUTION OF BUSINESS BIRTHS AND DEATHS TO OVERALL EMPLOYMENT MOVEMENTS

Patricia Getz, Jurgen Kropf, Sharon Strifas - Bureau of Labor Statistics
Patricia Getz, BLS, 2 Massachusetts Avenue, N.E., Room 4860, Washington, D.C. 20212

Key Words: Employment, Business Births, Time Series Modeling

NOTE: The opinions expressed in this paper are those of the authors and do not constitute official policy of the Bureau of Labor Statistics.

Background - One of the Nation's most widely watched economic indicators is the monthly estimate of nonfarm payroll employment from the Current Employment Statistics (CES) program, a sample survey of nearly 400,000 business establishments. Due to the dynamic nature of the U.S. economy, business births and deaths can be important contributors to the overall month-to-month movements in nonfarm employment. Yet throughout the 50 year history of the CES program, the measurement of employment change resulting from these components has been problematic. Part of this difficulty arises from the complication and expense of maintaining an up-to-date sampling frame for a continually changing universe of business establishments, within a monthly survey environment. In addition, the true magnitude of the birth/death issue has always been difficult to assess because it was confounded by other program limitations, especially the lack of a standard probability-based sample design. CES is now undergoing a comprehensive survey redesign that will place it on a probability basis and include quarterly sample frame updates to record births and deaths as quickly as possible on the frame; however, the sampling frame which is based on administrative records from the state unemployment insurance (UI) tax system is always a minimum of 6 to 9 months out of date with respect to the current reference month for CES estimates. For example, when CES employment estimates are being generated for September 1997, UI universe data for first quarter, through March 1997 is the most recent frame available. This UI universe sampling frame also serves as an after-the-fact benchmark and quality measure for the CES employment estimates for the month of March each year. As such, employment levels which will eventually be reported in the March UI figures must be accounted for in the CES estimates.

There are two basic approaches being considered by BLS to account for the employment due to births and deaths in the period before they can be reflected on the standard sampling frame: a model-based approach using time series techniques and a sample-based approach that is dependent on constructing an auxiliary, and more timely sampling frame for business births. This paper discusses some of the research to date on a model-based approach.

Profiling Sources of Employment Change - In order to be a viable candidate for time series modeling, a data series must exhibit relatively predictable behavior over time. In addition, in the CES application it was deemed important that the birth and death components be sufficiently small contributors to overall employment change so that the heavy majority of the change was not being measured through birth/death modeling, but through the sample survey. If this were not true, model-based estimation for this component might not be considered viable in the redesigned survey.

As part of an initial exploratory data analysis the contribution of each of the three major components to employment change were tabulated both at state and national levels from 5-year longitudinal files of UI universe microdata:

- Each individual UI account was classified as a birth, death, or continuing unit for each 1-year span, March 1991 to March 1995, according to the following definitions:

  • Births , employment in March $t = 0$ and March $t+1 > 0$;

  • Deaths, employment in March $t > 0$ and March $t+1 = 0$; and

  • Continuing Units, employment in March $t > 0$ and March $t+1 > 0$.

Table 1 summarizes overall findings at a US total level.

- Births and deaths are each quite large but also offset each other to a great extent, such that the net/birth is relatively small as a percent of over-the-year change overall. For example in the 1994-5 period, overall employment change is 3.3 million with a 3.0 million increase in the continuing unit population. The births add 3.9 million but are offset by a subtraction for deaths of 3.6 million. Thus the net birth/death contribution is 0.3 million.

- However while the net birth/death is small in general, the importance of these components can vary by industry, geography, or time period. For example,

in 1991-2, the only year of employment decline during the study period, most of the over-the-year decline came from deaths. The continuing units over-the-year employment change is near zero while the net of births and deaths is -357,000 compared with an overall over-the-year change of -363,000.

TABLE 1
Over the Year Employment Change
March to March
Total Private Employment
(BLS Universe Database)

	Overall Change	Continuing Units	Birth Units	Death Units	Net Birth/ Death
YEAR					
1991-92	-363,000	-5,000	3,971,000	-4,330,000	-357,000
1992-93	1,463,000	1,449,000	3,890,000	-3,876,000	14,000
1993-94	2,740,000	2,663,000	3,754,000	-3,697,000	58,000
1994-95	3,308,000	2,977,000	3,944,000	-3,613,000	332,000

Focusing on a six state subset of the national data, further exploratory data analysis was completed for over-the-month as well as over-the-year change. The states used were: California, Florida, Michigan, New Jersey, New York, and Pennsylvania. Results were analogous to the over-the-year changes:

- the net of births and deaths was small although each component is relatively large each month,
- the net of births and deaths was small relative to the continuing unit change component.

Another important result from the exploratory data analysis was that when viewed over the entire 5 years of the study period: the net birth/death component appeared to have some cyclical sensitivity -- it was negative during the earlier years in the study (which coincide with negative or flat employment growth economy-wide) and then gradually accelerated over time, correlating with the overall stronger employment growth observed in those years and yielding a net positive in the last two years observed.

This exploratory data analysis indicated that modeling for the net birth and death component of employment change may be a feasible option. The net appears sufficiently small and correlated with both its own past history and overall employment trends that can be observed from the continuing unit population via the ongoing monthly sample, that simple time series models may be appropriate for estimation.

Time Series Models for Birth and Death Employment - Although most users of CES employment data are interested in the over-the-month change, the CES survey actually produces and publishes employment levels on a monthly basis; the change is calculated simply as the difference between two levels. Thus, under a birth/death modeling scenario, the total CES employment level estimate for a given month can be described as the sum of continuing units employment (estimated from the sample) plus the net of birth and death employment for the given reference month plus the continuing employment from births born in previous months.

Net birth/death models - Two similar modeling techniques were initially examined: regression and ARIMA.

Regression models - To predict birth and death employment in a current month, models for individual States were fit using a variety of lagged values of births and deaths as the independent variables. Additionally, there was some experimentation with using trends from the continuing unit population because the exploratory data analysis had shown the birth/death movements exhibiting some cyclical sensitivity; however, this did not improve the model fitting and it was dropped from this round of testing. A backward elimination of other insignificant variables (i.e., lagged birth and death values) was done to arrive at the best model fit. Models were initially fit using the UI universe data for January 1990 through March 1994 and forecasting capabilities tested for the January through December 1995 period, assuming a 9-12 month forecast horizon. The first forecast then was for January through March 1995 using data through March 1994 in the modeling. The final forecasted quarter, October through December 1995 used data through December 1994 in the model fitting.

Based on the R-squared, F-test, and t-test values, the model fits were generally good; however, the forecast errors are fairly large in percentage terms, ranging up to 20% when measured as mean errors.

ARIMA models - Fitting of ARIMA models to these birth, death, and net birth/death series yielded similar results to the regression modeling, as expected. Initially models were fit for individual state series by proceeding through the routine steps of ARIMA modeling: data differencing and transformations, review of ACF and PACF to select the AR and MA parameters to be tested, and review of model fit diagnostics including AIC coefficients, p-values on the white noise test and check of the residual plots. Forecast errors for the best-fitting models for each state series were

comparable to those achieved under regression model forecasting. The best-fitting models were within the class of commonly used ARIMA models. This was an encouraging result because, as noted above, models will be needed for many geographic and industry level series in actual CES production. After manually fitting the models, fitting through the X-12 automatic routine was tested adding just one other model, a (310) (011) to the 5 already included in the default routine; the five are: (011)(011), (012)(011), (210)(011), (022)(011), (111)(011).

Forecast errors achieved under this procedure were comparable to those mentioned above for the regression models.

Second Round Model Specification - Combining the best results from the first round testing with a search for a method that could prove operationally feasible for the hundreds of industry and geographic levels that CES produces estimates for lead to testing the Census Bureau's REGARIMA (regression with autocorrelated error terms) software, including the automatic model selection feature. REGARIMA models were specified for a single series that is the net of employment from births, deaths, and continuing births, rather that the three series individually. This represents the net of what is not covered by the continuing units sample, but still must be accounted for by the CES estimates. An independent regression variable defined as over-the-month employment change in the continuing unit population was added to the simple ARIMA models. This draws on the correlation observed between continuing units behavior and the behavior of the net/birth death series in the exploratory data analysis, and adds a component of cyclical sensitivity. This model specification assumes births and deaths are not measured by the CES sample in the month they occur and estimates a net of the two for each reference month. In addition, it estimates current employment levels for birth units from the previous 11 months as these are not yet represented in the sample-based estimates either.

Modeling Results - To date the REGARIMA modeling has been tested for all 2-digit SIC level series for the national CES data series. Forecast errors were quite large in percentage terms for some of the smaller industries, but largely offset at higher level aggregates such that a very small error was achieved for the total private employment level, the most important figure from the CES survey. The average percentage error for total private for the 12 months that the forecasts were tested was 0.55% (table 2).

For most of the 70 2-digit SIC level series that were separately modeled, the REGARIMA model using over-the-month change in continuing units worked best, but for a few a simple ARIMA model with no regression variable gave better diagnostics and forecasts and was chosen as the best model. Forecasts from each of the best models at the 2-digit level were used to derive the one-digit SIC level and total private series shown on Table 2.

State Level Modeling - In addition testing has begun for REGARIMA modeling at State major industry division levels. Results for the first industry tested, wholesale trade indicated that the sum of State level forecasts were very close to independent forecasts done at the national level. This is an important characteristic because there should be consistency between national and State published CES series. Because CES is a federal/state cooperative program, BLS independently generates national level estimates while each state employment security agency produces and publishes state and metropolitan area estimates.

Further Research - Data limitations have allowed for forecast testing for only one year thus far. A second year of historical universe data will soon be available and usable for further forecast testing. In addition more extensive model testing for state level major industry division and 2-digit SIC series will be conducted. Finally the model testing will be extended to metropolitan area series. Any modeling techniques adopted for CES must be robust enough to support these levels of detail.

Modeling results also must be evaluated against their contribution to CES benchmark revisions, i.e., the difference between the CES estimate for March of each year and the final benchmark level which established form the UI universe levels. The ultimate test of whether the forecasts are 'good enough' lies in whether they help bring CES estimates to within historical levels of benchmark revisions, at a minimum.

TABLE 2

Results of "Best Models" for Birth Estimates

"Best Model" is determined by significance of independent variable (t-statistic)

	Jan-95	Feb-95	Mar-95	Apr-95	May-95	Jun-95	Jul-95	Aug-95	Sep-95	Oct-95	Nov-95	Dec-95	Avg. Diff	% Diff
Industry														
Total Private														
Actual Net-Birth	2592168	2857723	2952173	3127869	3022109	3105652	3211440	3184535	3232910	3366042	3285048	3303553		
Birth Estimate	2275458	2874657	2942053	3074191	3088154	3153289	3170208	3229015	3208856	3232405	3262381	3524125		
Difference	316710	-16934	10120	53678	-66045	-47637	41232	-44480	24054	133637	22667	-220572	17203	0.55
Mining														
Actual Net-Birth	4945	10668	11124	12598	12759	13269	12293	13470	13857	13850	14279	13668		
Birth Estimate	9876	14243	14855	15328	15145	13565	15277	13951	14578	15708	16152	17927		
Difference	-4931	-3575	-3731	-2730	-2386	-296	-2984	-481	-721	-1858	-1873	-4259	-2485	-20.32
Construction														
Actual Net-Birth	224809	231353	244169	276445	266207	281379	298014	291403	296858	318697	300161	297362		
Birth Estimate	229012	223624	236634	265464	263722	275947	298089	290610	291444	311266	294621	303478		
Difference	-4203	7729	7535	10981	2485	5432	-75	793	5414	7431	5540	-6116	3579	1.29
Manufacturing														
Actual Net-Birth	225774	337973	347050	348024	337878	343899	331643	363539	366243	354962	354204	351544		
Birth Estimate	282417	338002	344951	350132	357440	355117	342098	346177	346425	332572	339530	372327		
Difference	-56643	-29	2099	-2108	-19562	-11218	-10455	17362	19818	22390	14674	-20783	-3705	-1.09
TPU														
Actual Net-Birth	107950	142949	149153	154428	153820	157190	163257	162034	167351	172120	166965	169394		
Birth Estimate	91338	118471	121566	122591	124497	128628	117248	128409	128141	130434	128256	139880		
Difference	16612	24478	27587	31837	29323	28562	46009	33625	39210	41686	38709	29514	32263	20.74
Wholesale														
Actual Net-Birth	133523	154373	156471	161302	156077	156710	159228	160410	160067	163669	165903	166706		
Birth Estimate	85545	146114	146995	143819	151623	154670	135798	155482	151032	145961	155270	183863		
Difference	47978	8259	9476	17483	4454	2040	23430	4928	9035	17708	10633	-17157	11522	7.30
Retail														
Actual Net-Birth	700323	691975	714262	768373	742969	762210	794270	774553	788238	836984	822545	831182		
Birth Estimate	628993	745503	767168	795233	805764	820977	825194	848544	857114	881331	897748	947575		
Difference	71330	-53528	-52906	-26860	-62795	-58767	-30924	-73991	-68876	-44347	-75203	-116393	-49438	-6.43
FIRE														
Actual Net-Birth	124385	177369	181302	185173	176324	181432	177601	185557	186916	195206	191678	194990		
Birth Estimate	128799	146958	151299	148497	154818	160272	158608	161672	165674	168660	171759	183759		
Difference	-4414	30411	30003	36676	21506	21160	18993	23885	21242	26546	19919	11231	21430	11.92
Services														
Actual Net-Birth	1070459	1111063	1148642	1221526	1176075	1209563	1275134	1233569	1253380	1310554	1269313	1278707		
Birth Estimate	819478	1141743	1158584	1233128	1215145	1244114	1277896	1284171	1254448	1246473	1259046	1375316		
Difference	250981	-30680	-9942	-11602	-39070	-34551	-2762	-50602	-1068	64081	10267	-96609	4037	0.33