IMPROVED TIMELINESS OF GEOGRAPHIC DATA THROUGH AUTOMATION OF THE CURRENT EMPLOYMENT STATISTICS PROGRAM

Kenneth Shipp, Bureau of Labor Statistics
2 Massachusetts Ave. N.W., Washington, D.C. 20212

"It takes so long to create some statistics that by the time they are published, they no longer accurately reflect reality;
New Technologies could increase the timeliness of government statistics and reduce the burden on firms." -- Kelly and Wyckoff

"A stitch in time saves nine" -- Mother Goose

The demand for economic data on a timely basis has never been greater. Government policy makers and businesses need current information to make decisions affecting interest rates, taxes, investment, and plant locations. Reliance on old methods of collection and delivery are not sufficient in an information age where relevant punctual data is needed to make better government policies and better private sector decisions. This paper describes how the Bureau of Labor Statistics and State Employment Security Agencies have responded to improve the timeliness and release of geographic economic data using new technologies. The release of comprehensive state and area employment data were published 10 weeks after the collection period. This recent BLS initiative reduced that time to 7 weeks.

The Current Employment Statistics Program

The Current Employment Statistics (CES) Program provides for the monthly collection and publication of detailed industry data on employment, hours, and earnings at the national, State and area levels. Based on a voluntary sample of nearly 400,000 nonagricultural establishments, it is the largest monthly survey of its type in the social science field.

The survey has provided economists with important time series of data since its inception in 1915. The CES survey serves as a cornerstone times series for economists, business people and policy makers tracking industrial growth and geographic movements of industry across the country. The survey currently releases 3,500 national and 22,000 State and area time series.

The CES estimates are particularly valuable because they are among the earliest available economic indicators each month. National estimates are available within 3 weeks of the collection period. The CES collects information about all employees, women workers, production workers, and hours, earnings and overtime hours for these production workers. The publication of estimates has immediate impact for policy makers, financial markets, and for business decision makers.

Estimates of all employees and aggregate hours serve as major sources for the calculation of the Gross National Product and the Federal Reserve Board's Index of Industrial Production. The manufacturing workweek and average overtime hours are among the most sensitive of the leading economic indicators. The average hourly earnings series provide a measure of labor costs to employers and are used to escalate labor agreements. Changes in the CES data lead to important economic decisions. For example, based on changes in the trend of these series, tax regulations are instituted, revoked or modified, the money supply is adjusted, interest rates are changed, collective bargaining decisions are determined, and decisions about plant locations and investment are made.

BLS and State Partnership

The CES program operates as a Federal/State partnership. BLS acts as the manager of the program and the State Employment Security Agency within each state performs much of the work. Through a contractual agreement, BLS provides funding, sets program goals, objectives, and schedules, determines methodology and gives technical guidance to the States to assure that statistical standards are followed. Although independent from BLS, each State agency works within these BLS guidelines to conduct the CES survey.

Individual state employment security agencies are responsible for data collection, analysis, and dissemination of their own State and area estimates. State agencies carry out basic labor market information work, utilizing BLS technical directives. In addition, they respond to the data needs associated with the policies of the U.S. Department of Labor programs. Another important State function is the transmission of data to BLS. While each state is responsible for its own data, BLS acts as the focal point for interstate comparisons--collecting and disseminating labor

market information for all states, the District of Columbia, and over 275 major metropolitan areas. BLS also consolidates the data collection efforts of these individual states to produce independent national estimates from the sample collected by the states. This coordination and cooperation of Federal and State agencies allows the same sample to be used for computing national industry estimates in Washington by BLS and for computing individual State and metropolitan industry estimates in each State.

Timing--Collection and Publication of Data

Data collection activities within the CES survey make up a substantial portion of the survey's workload. The CES survey collection period uses the standard establishment collection period, the pay period which includes the 12th of the month. Between the 12th day of the month and the last Friday of the month, state agencies are responsible for the initial collection of data from the establishment, for non-response prompting, for editing outliers, and for providing that microdata to BLS in Washington. BLS then uses this microdata to produce estimates for the nation at detailed industry levels. [1] Of the 400,000 survey respondents, this initial collection cycle typically has a 55-60% response rate. (See the state by state response rates for 1983 vs. 1994 for the initial national estimate in table A.) The timing of this initial collection cycle is set to provide sample data for the first estimate of national data about 3 weeks after the reference week. For example, August 1995 estimates are published September 2. The full collection cycle for national estimates includes the next two months in which final response rates approach 95 percent.

---

[1]*Caution in aggregating State data.* The national estimation procedures used by BLS are designed to produce accurate national data by detailed industry; correspondingly the State estimation procedures are designed to produce accurate data for each individual State. State estimates are not forced to sum to national totals nor vice versa. Because each State series is subject to larger sampling and nonsampling errors than the national series, summing them cumulates individual State level errors and can cause distortions at an aggregate level. This has been a particular problem at turning points in the U.S. economy, when the majority of the individual State errors tend to be in the same direction. Due to these statistical limitations, the Bureau does not compile or publish a "sum-of-States" employment series. Additionally, BLS cautions users that such a series is subject to a relatively large and volatile error structure, particularly at turning points in the economy.

Table A. Sample Available:
Initial National Estimates

March 1983				March 1994		
Sample Units	Sample Employ - ment	Response Rate (%)	State	Sample Units	Sample Employ - ment	Response Rate (%)
1,357	202,397	45	AL	4,691	544,656	55
13	349	1	AK	809	78,807	58
1,608	263,289	54	AZ	2,296	457,311	47
994	99,948	53	AR	2,793	302,664	69
8,309	1,043,587	37	CA	18,813	2,179,542	51
901	141,226	35	CO	3,033	375,214	53
1,506	238,270	48	CT	3,016	371,461	63
342	18,319	21	DE	1,144	80,708	45
231	24,455	40	DC	592	33,759	43
3,282	321,336	48	FL	10,778	1,065,556	58
1,154	194,931	44	GA	6,119	709,466	64
878	107,653	56	HI	1,039	186,752	67
648	37,944	40	ID	2,448	104,477	57
5,361	487,737	51	IL	9,678	964,492	52
1,463	312,948	45	IN	3,343	425,549	43
1,590	187,828	59	IA	3,642	293,919	60
1,663	162,524	47	KS	3,505	230,934	64
754	131,427	46	KY	2,551	358,666	56
939	145,190	45	LA	3,200	322,825	56
1,175	113,527	63	ME	1,707	166,004	77
1,985	207,881	42	MD	3,785	484,447	62
1,849	274,186	42	MA	5,672	682,317	55
1,809	325,749	32	MI	6,928	840,016	51
1,245	145,183	42	MN	3,169	393,571	55
1,354	127,388	42	MS	3,057	319,597	62
2,111	229,346	49	MO	4,453	580,286	68
365	20,678	35	MT	1,272	69,041	54
1,042	73,763	49	NE	2,501	203,191	67
630	68,502	48	NV	1,317	164,382	51
426	53,774	51	NH	1,328	118,286	68
2,923	282,046	42	NJ	6,751	772,797	60
686	51,604	53	NM	1,995	162,720	48
8,961	979,543	44	NY	12,151	1,578,498	55
2,578	379,636	48	NC	6,165	835,300	46
1,034	44,142	40	ND	1,305	45,298	41
3,899	591,237	49	OH	6,018	933,635	54
1,033	122,214	48	OK	2,349	205,228	55
1,051	81,873	26	OR	2,625	249,957	41
7,400	645,778	54	PA	10,555	1,091,021	65
747	103,995	68	RI	975	99,476	70
2,675	335,346	68	SC	3,466	503,484	71
619	33,618	49	SD	1,189	77,735	52
1,111	212,418	45	TN	6,625	638,003	66
3,401	406,936	36	TX	11,361	1,253,826	46
655	64,164	48	UT	2,024	173,578	55
785	36,534	58	VT	1,314	82,793	77
3,504	433,382	51	VA	6,528	698,881	53
914	106,944	30	WA	3,734	504,761	54
799	97,934	47	WV	2,932	207,483	56
1,677	242,879	48	WI	4,250	676,524	65
262	13,711	34	WY	1,605	53,736	44
93,698	11,027,269	45	US	214,596	23,952,630	55

Before this initiative, individual state and area estimates typically were produced by each state two to three weeks later than the national due dates using approximately 80% response rates. These estimates are published by each State Employment Security Agencies in individual state press releases and are also forwarded to BLS in Washington. BLS then consolidates these individual State and area estimates into a comprehensive geographic press release which includes CES employment data for major industry sectors for all states. Comprehensive industrial data for all states and most metropolitan areas are also provided to the user community through electronic data bases at the same time the press release is published. These comprehensive data, 22,000 time series, include estimates for employment at the three digit Standard Industrial Classification (SIC) level as well as hours and earnings of production workers in manufacturing. In 1994 BLS and State Employment Security Agency's enhanced the press release by publishing seasonally adjusted statewide employment series for major industry divisions.

BLS started publishing the “State and Metropolitan Area Employment and Unemployment” press release in 1983 when many of the local labor market analysis functions were transferred to BLS from another agency within the Labor Department. At that time the publication schedule for the geographic data was considerably lagged in relation to the national estimates. The national estimates are published *in The Employment Situation* press release only 3 weeks after the reference week. While geographic estimates were unavailable until 10 weeks after the reference week. Much of this lag in publishing geographic data was because of a heavy reliance on mail and on unique state operational structures which made it difficult to standardize operations and control production timing. In 1983, the sample (215,000) was solicited, collected, and prompted exclusively by mail. Each state had a unique computer system which made standardization of methodologies and operations difficult to control. And the expanded use of telecommunication to collect and transmit data was in its infancy.

Operational Improvements

Between 1983 and 1994, BLS and State Employment Security Agency's successfully implemented major operational and technological improvements in **data collection**, **systems standardization**, and **telecommunications** which allowed the publication schedule for geographic estimates to be reduced by 3 weeks, without adversely effecting the estimates.

Data collection of CES reporters today is remarkably different from the mailing scenarios of the mid-1980s. A decade of research and implementation of Computer Assisted Telephone Interviews (CATI), Touch-tone Data Entry (TDE), Voice Recognition (VR), and FAX technologies have transformed the collection of sample data. These new technologies now

collect nearly half of the total sample employment (48,000,000 employees). These new methods have substantially reduced the time previously required to solicit, receive, and prompt reporters through the mail. State response rates for the preliminary national estimates in 1983 were about 45% of the sample. Overall response rates are now approaching 70% in many states for this initial estimate, with the automated CATI and TDE methodologies leading the way by consistently having 75-85% response rates. These advances have been made as total sample has grown from 215,000 to 395,000 in this time frame. And, sample for the initial estimate has grown from 95,000 to 225,000. The data collection efforts have reduced mail costs and improved the initial national employment estimate's reliability. Between 1983 and 1994 the preliminary to final estimate revision at the aggregate total non farm level was reduced from 85,000 to 68,000 employees. **Werking, Clayton, et. al.**

Table B: Data Collection Improvements:
Initial National Estimates

	March 1983		
Collection Method	Sample Units	Sample Employment	Response Rate (%)
CATI	0	0	NA
TDE	0	0	NA
MAIL	93,698	11,027,269	45
ALL	93,698	11,027,269	45
	March 1994		
	Sample Units	Sample Employment	Response Rate (%)
CATI	12,083	3,217,227	84
TDE	42,091	9,829,610	77
MAIL	160,422	10,905,793	50
ALL	214,596	23,952,630	55

Systems standardization across State Employment Security Agency's was also addressed by BLS during this time frame. In response to the Levitan Committee's recommendations for improving the CES, in 1986 BLS and State Employment Security Agency's developed and began exporting the Automated Current Employment System (ACES). This system has incorporated the automated data collection improvements (e.g. TDE) as well as standardizing other state CES operations such as solicitation, estimation, editing, benchmarking, and transfers of files. ACES is now operating in 38 states, the District of Columbia and Puerto Rico. ACES provides BLS with

- standardization of processes;
- cost efficiency in system development, training and maintenance;
- automation of previously manual activities;
- file transfers and data interchange with other necessary data sources such as the Unemployment Insurance data files necessary for sample solicitation, and benchmarking.

ACES also facilitates cross state processing by making exchange of multi-unit reporters and other shared sample data (used in interstate metropolitan area estimates) more timely and reliable. It automatically transmits sample reports, sample control information, and geographic estimates through standard telecommunications files and protocols. These standard files and coordination with other BLS systems have helped streamline the monthly publication cycle. For example, each state using ACES will transmit a comprehensive set of geographic estimates to a BLS computer system that detects and creates a file of potential outliers for further review by BLS regional staff and state analysts. These outliers are reviewed, corrected, reprocessed, and updated on public data bases within one day. Before these ACES and telecommunications enhancements, this process was dependent on mail to receive, review, and revise the estimates. **Shipp.**

Improvements in ***telecommunications*** by BLS and State Employment Security Agencies have played a major role in reducing the time frame for producing CES estimates. As early as 1979 BLS began using telecommunications of sample microdata from State Employment Security Agencies to improve the response rates for the initial national estimates. Now, BLS and State Employment Security Agencies rely on telecommunications for quick transfers of sample data, sample control information, edits, and geographic estimates. Each state is linked through high speed modems to central processing sites of BLS as well as to other states. Entire processes have been influenced by file transfers through telecommunications. For example, each month, states produce and transmit a file of delinquent reporters to 3 major data collection centers staffed by BLS. These centers prompt these delinquent reporters through automated non-response lists and faxes. Reporters then respond through automated TDE software into ACES. These reports are edited and transmitted to Washington for inclusion in the national estimates and used in subsequent State and area estimates.

Many other CES operational improvements have been influenced by telecommunications. Transmission of sample data from states can now be done on a flow basis, with many reports arriving as late as Monday of the publication week. These reports are available to national BLS staff using sophisticated visual screening technologies running on an independent Local Area Network. **Esposito, et. al.**

These technological enhancements were primarily developed in response to improving the reliability of initial national estimates. However, as they became successfully implemented, the opportunity to improve the timeliness and accuracy of geographic estimates was also available.

Relating Improvements to CES Geographic Data

***Faster publication*.** BLS has long encouraged State Employment Security Agencies to produce geographic CES estimates in conjunction with the release of national CES data. However, very few states have been able to accomplish this goal because of resource and timing conflicts between BLS/State Employment Security Agencies workload requirements. BLS sets minimum sample levels in contractual agreements with State Employment Security Agency's to obtain in the initial national CES estimate, which dictates that State Employment Security Agency's dedicate the last week in each month to the sample collection and editing process rather than in their individual states' estimation and analysis activities. Knowing these constraints, BLS revised the concurrent publication goal. To reduce conflict with national collection priorities, the revised goal was set to publish the "State and Metropolitan Area Employment and Unemployment" press release within 7 weeks of the reference week--a 50% reduction in the lag time between release of national CES data and geographic estimates. In February 1993 BLS proposed that State Employment Security Agency's revise their current individual publication cycles to produce geographic estimates four weeks after the reference week or about one week after the national CES estimates were published. Test simulations to measure the effect of this revised schedule on estimates and workload were conducted in 4 states who volunteered to run two sets of February 1993 estimates--a normal timing cycle and one test cycle two weeks before the normal cycle.

Total nonfarm employment--February 1993

	early	normal
Missouri	2,216,600	2,216,100
Montana	316,900	316,200
New Mexico	597,200	593,600

* Iowa also participated in the test, with similar results, but results were discounted because of operational problems.

Although these simulations show small differences at aggregate levels, states voiced concern about detailed SIC levels, particularly in those industries where reporters were on monthly payrolls and would not have adequate representation in these early estimates. BLS agreed with the states concerns about the collection cycle and restored one of the two State Employment Security Agencies collection weeks that was cut in the initial proposal. These revised time frames were acceptable within both State Employment Security Agencies and BLS operations. In March 1994, BLS and State Employment Security Agency's implemented the revised schedule. The goal for faster publication of the "State and Metropolitan Area Employment and Unemployment" release was therefore achieved by reducing the State Employment Security Agencies collection cycle by one week and also through reducing one week required in the manual publication process in Washington. For an ever increasing number of computer users of the BLS electronic database at the National Institutes of Health (NIH), the gains were even greater. Electronic access through BLS NIH databases to comprehensive geographic data (22,000 time series) was reduced to 6 weeks after the sample reference week.

Accuracy. Often, gains in timeliness come at the expense of accuracy or coverage. With a year of this revised schedule completed, we can determine if these gains in timeliness came at the expense of accuracy by comparing revisions between preliminary and final estimates at the Statewide level. Since the complete scope of States, areas, and industries remained the same in both publication cycles, there has been no change in coverage or scope. A comparison of statewide total nonfarm employment revisions for February-November in 1993 (old schedule) and February-November 1994 (revised schedule) shows relatively small changes for any individual state and practically no change when evaluating revisions across all states (See Table C: "Revisions in Preliminary to Final Employment Estimates").

Average statewide absolute preliminary to final change (old schedule) 0.4

Average statewide preliminary to final change (new schedule) 0.4

Table C. Revisions in Preliminary to Final Employment Estimates: Total Nonfarm

Average Monthly Revision (in 000's)

State	Old Schedule--1993	New Schedule--1994	Employment
AL	0.4	-1.7	1,752.5
AK	0.1	0.1	260.1
AZ	0.6	0.0	1,685.2
AR	0.2	0.4	1,035.2
CA	0.1	-1.5	12,136.1
CO	-0.1	0.6	1,749.7
CT	-0.4	-0.3	1,542.4
DE	-0.1	0.2	355.0
DC	-0.5	0.2	657.3
FL	1.3	1.8	5,796.6
GA	0.3	0.3	3,263.8
HI	-0.1	0.0	536.1
ID	0.3	0.2	463.0
IL	-0.9	3.8	5,463.1
IN	-0.2	0.2	2,712.0
IA	-0.2	0.4	1,319.2
KS	0.1	0.1	1,166.3
KY	0.5	0.7	1,598.7
LA	0.9	1.4	1,727.1
ME	-0.1	0.1	531.2
MD	0.2	-1.0	2,144.5
MA	0.8	-0.1	2,905.0
MI	-0.8	-1.1	4,141.6
MN	0.8	-0.3	2,311.4
MS	-0.9	0.8	1,053.4
MO	0.9	1.3	2,472.9
MT	0.3	0.1	340.5
NE	-0.5	0.2	795.5
NV	0.4	0.4	736.7
NH	0.5	0.2	522.3
NJ	0.4	-0.4	3,550.3
NM	0.5	0.5	658.1
NY	3.1	0.7	7,800.3
NC	2.9	2.8	3,361.1
ND	0.2	0.2	294.7
OH	-0.6	0.3	5,076.2
OK	0.9	1.2	1,278.6
OR	0.0	0.2	1,364.0
PA	0.9	2.0	5,187.8
RI	0.0	-0.4	434.0
SC	-0.3	-0.3	1,607.3
SD	0.3	0.4	332.9
TN	1.1	0.9	2,420.8
TX	3.9	2.1	7,740.1
UT	0.6	-0.2	861.0
VT	-0.1	0.2	263.8
VA	1.9	-0.6	3,006.1
WA	1.1	1.6	2,309.0
WV	-0.2	0.2	674.8
WI	1.1	1.0	2,482.5
WY	0.4	-0.1	216.9

The implementation of telecommunications and data collection have been instrumental in providing states with the necessary sample to publish data faster without adversely effecting accuracy.

Standardization of state operations through software like ACES has also raised states' productivity and ability to produce better data. In fact, the 1994 preliminary State estimate is made with substantially more sample than the 1983 state estimate.
We will continue to monitor the preliminary to final revisions in these data in light of this project as well as other BLS initiatives to provide more timely geographic data.

Future enhancements. BLS and State Employment Security Agency's are committed to providing even more timely geographic CES data to users. In the short run and within the current panel sample methodology there are incremental changes that can be made to provide data to users somewhat faster. We have identified two days in the production cycle that could be reduced by streamlining the editing of macro estimates. There may be other time savings available by automating some of the manual activities now inherent in the production of the paper documents.. These short run opportunities could make the data available on databases within 5 weeks of the reference week, and in printed form within 6 weeks. Much of our current focus in meeting users data needs has turned to electronic formats and accessibility of data. The explosion of Internet gives BLS an opportunity to provide timely comprehensive data to an expanded base of users at minimal costs.

In the longer run, the biggest advances will likely be results of a more comprehensive redesign of the CES survey that is beginning in 1995. This redesign will transform the CES into a probability design that will likely accommodate state, area, urban and rural components, i.e. geographic data are available to determine *where* industrial employment, hours and earnings movements are taking place at the same time national data is available. These advances are dependent on the continued success of automated data collection methods, systems advancements, and especially telecommunications....as opportunities like Electronic Data Interchange (EDI) are available for data exchange between respondents, survey takers, and data users.

And, in any time frame we will continue to be conscious of users needs--as a general rule, users want extensive coverage; ease of access; timeliness, accuracy (validity, reliability); clear-cut methodology (consistency and continuity); and integrity of both the data and the source. When the producers of data keep these requisites in mind, the result should be strong demand for the figures. **Duncan, Gross.**

REFERENCES

Clayton, R.L. and L.J. Harrell, Jr. (1989), "Developing a Cost model of Alternative Data collection methods: MAIL, CATI, and TDE," *Proceedings of the Section on Survey Research Methods, American Statistical Association,* pp. 264-269.

Clayton, R.L., R.J. Rosen, and T.B. Robino, Jr., (1993), "Non response Prompting Behavior an Establishment Survey," *Proceedings of the Section on Survey Research Methods, American Statistical Association,* pp. 745-748.

Duncan, Joseph W. and Andrew C. Gross (1993), *Statistics for the 21st Century*, New York: Dun and Bradstreet. p. 75.

Esposito, Richard, Dong-yow Lin, and Kevin Tidemann, "The ARIES Review System in the BLS Current Employment Statistics Program", *ICES--Proceedings of the International Conference on Establishment Surveys*, June 1993.

Kelly, Henry and Andrew Wyckoff, "Distorted Image," *Technology Review*, February/March 1989, pp. 53-60.

National Commission on Employment and Unemployment Statistics (1979), *Counting the Labor Force*, Washington, U.S. Government Printing Office, p. 162.

Scott, Stuart, George Stamas, Thomas J. Sullivan, and Paul Chester (1994), "Seasonal Adjustment of Hybrid Economic Time Series." *Proceedings of the Section on Business and Economic Statistics, American Statistical Association.*

Shipp, Kenneth (1991), "Cooperation of Federal/State Agencies in the Current Employment Survey: A Case History of Automation." *Proceedings of the Section on Business and Economic Statistics, American Statistical Association,* pp. 152-156.

Shipp, Kenneth and Tom Sullivan (1992), "Using X-11 ARIMA to Seasonally Adjust State Level Industry Employment Data in the Current Employment Statistics Program." *Proceedings of the Section on Business and Economic Statistics, American Statistical Association,* pp. 349-354.

U.S. Department of Labor (monthly), *Employment and Earnings*, Washington, U.S. Government Printing Office, technical note.

Werking, George S. and R.L. Clayton (1994), "Integrating CASIC into the Current Employment Statistics Survey," *Proceedings of the 1994 Annual Research Conference and CASIC Technologies Interchange,* pp. 738-748.

Werking, George S., A.R. Tupek, and R.L. Clayton (1988), "CATI and Touch-tone Self-Response Applications for Establishment Surveys," *Journal of Official Statistics.* Vol. 4. No. 4. 1988. pp. 349-362.

Winter, Debbie L.S. and R. L. Clayton (1990), "Speech Data Entry: Results of the First Test for Voice Recognition for Data Collection," *Proceedings of the Section on Survey Research Methods, American Statistical Association,* pp. 387-389.