

Challenges in Measuring and Disseminating Employment and Unemployment Estimates

– The U.S. Experience

Erica L. Groshen

Bureau of Labor Statistics, Washington DC, USA– Groshen.Erica@bls.gov

Anne Polivka*

Bureau of Labor Statistics, Washington DC, USA – Polivka.Anne@bls.gov

Abstract

In a world with dynamic markets and technological changes, statistical programs will always face new challenges and the need to adapt practices. As one of the longest running statistical programs, the US employment and unemployment estimates provide a rich case study of the many ways statistical program may need to adapt. This paper reviews some ongoing and new challenges faced by BLS in producing estimates presented in the Employment Situation and steps undertaken to address them with regards to sampling, cost of data collection, response rates, maintenance of the comparability of estimates over time, and increasing demands for new or different information about the labor market.

Keywords: Employment; unemployment; Current Population Survey; Current Employment Statistics.

1. Introduction Every month the Bureau of Labor Statistics issues a press release—called the Employment Situation—reporting monthly estimates of the unemployment rate, employment-to-population ratio and the labor force participation rate for the U.S. as a whole and various demographic groups. Also in the release are estimates of the total number of jobs and changes in the number of jobs from the previous month for the entire U.S. economy and by detailed industry groupings. The unemployment estimates and other labor force estimates for specific demographic groups are derived from the Current Population Survey (CPS) - a household survey conducted for BLS by the Census Bureau. The estimates of the number of jobs and monthly changes are derived from BLS's Current Employment Statistics (CES) - an establishment survey. Both of these surveys were in existence in the 1940s and have proven to be remarkably resilient over time. However, while providing new opportunities, recent changes in the social and technological environment in the U.S. also are posing challenges to the measurement of these statistics about the labor market.

2. Background on the CPS and the CES

Both the CPS and the CES are large, monthly sample-based surveys. However, their resemblance ends there. The CPS sample is an addressed based, multistage, stratified random sample of approximately 72,000 assigned housing units design to be representative nationally and for each geographic state. Surveyed housing units provide information for about 110,000 individuals monthly. To improve the precision of over the month and over the year comparisons, sampled households are interviewed for four consecutive months, not interviewed for the next eight months and are then interviewed again for four consecutive months. The CPS data is collected by interviewers using a computerized instrument. Households' first and fifth monthly interviews typically are conducted in respondents' homes, while the other monthly interviews are conducted over the phone either from interviewers' homes or from a centralized phone facility. The CPS is typically fielded the week containing the 19th of the month and ask questions about work activities during the week containing the 12th or prior 4 weeks for potentially unemployed job seekers. CPS estimates are released approximately 3 weeks after collection, usually on the first Friday of a month in the Employment Situation.

Every month, the CES obtains information from approximately 143,000 businesses and government agencies, representing approximately 588,000 individual worksites, about the size of their payroll and the hours and earnings of workers on these payrolls. For private employment, the CES sample is a stratified, random sample of establishments (worksites) clustered and drawn by employers Unemployment Insurance account numbers. Government employment estimates are obtained directly from payroll employment counts provide by government agencies. CES begins collecting data for the specified month as soon as an

establishment's pay period that contains the week of the 12th is completed. Initial preliminary monthly estimates typically are released in the monthly press release, typically the first Friday of the month. The collection period for the first, preliminary estimate varies between 9 and 15 days depending on when the press release is issued. Given the short time period and the fact that some establishments have payrolls that conclude at the end of a calendar month, many establishments are not able to provide their information in time to be included in the first estimates. Consequently, data from sampled units continue to be collected for 2 more months and estimates for the reference month are revised twice, with a second preliminary estimates being released the month following the initial release, and the final sampled based estimate being published two months after the initial estimate is released. CES data is obtained using a wide variety of data collection modes including Computerized Assisted Telephone Interviewing, Electronic Data Interchange, Fax, Web, and Touchtone Data Entry.

3. Sampling challenges and solutions

In deriving estimates from sample surveys, one key challenge is obtaining and maintaining the representativeness of the sample. The CPS challenge is to incorporate newly-constructed residences and incorporate shifts in the U.S. population. The CES challenge is to account for establishments born after the current sample was selected. Both surveys have undertaken measures to address these concerns.

Prior to the implementation of the 2010 sample redesign, the CPS sampled housing units from lists obtained from the most recent, once a decade, Decennial Population Census. Throughout the decade, between Censuses, the CPS sample was continuously updated using building permits and information from the Census Bureau's Housing Vacancy Survey to capture housing units built after the most recent Census. With the introduction of the 2010 redesign, annual sampling was introduced. Annual sampling became possible primarily because a Master Address File became available. The MAF is continually updated using several different procedures and sources including mailing addresses from the postal service. Compared to previous procedures, annual sampling permits the sample frame to be more up-to-date, provides better control of the survey sample size in response to changes in budget or data requirements, reduces the requirement for sample maintenance cuts if an area grows more than anticipated, and provides smaller sampling variance for many statistics of interest.

Starting in 2003, CES estimates were based completely on a probability sample. For its estimates, CES draws a new sample each year consisting of continuing units that were in the sample the previous year and newly drawn sample units (a third group- units that were in the previous year's sample, but are not in the current sample - generally are dropped from the sample). For newly selected business units, data collection specialists initiate and begin to enroll sampled units immediately after a sample is selected. However, to reduce costs the entire sample is not enrolled at once, rather sample units are enrolled throughout the year.

The collection of the survey employment data from a unit generally begins in the first month after it is initiated. At first, when probability sampling was introduced to the CES, data from newly initiated units were only introduced into CES estimates once a year. With this approach at least a full year would pass between when the sample was drawn, when the sample was fully implemented and when data from new units were introduced to the estimates. Given that units initiated early in the year may have been providing useful data for most of the preceding year, this procedure resulted in wasted data and caused the sample to be more dated than it needed to be; which in turn may have resulted in larger than need be benchmark revisions. In addition, the delayed inclusion of data from newly initiated units contributed to revisions in monthly estimates in November and December because preliminary estimates for these months did not include data from newly initiated units, whereas revised estimates did. To address these concerns, with the publication of the July 2014 first preliminary estimates, CES began introducing new sample units on a quarterly basis.

In addition to issues associated with the timing of the inclusion of data in CES estimates from newly sampled units, there also are issues with units not being included on the sampling frame due to new establishments opening since the frame was constructed. To address these concerns, CES models employment growth due to establishment births. The model procedure consists of two components. For the first component, units that

have stopped reporting are not excluded from the calculation of the current month's employment. Instead, previously reported employment for these non-reporting units are updated based on the month-to-month industry level estimates of change that were generated from the establishments that did report in consecutive months. This updating is done because research using frame data from 1995 to 2007 indicated that employment changes from establishment openings and closings essentially cancelled each other out. The second component involves modeling the small residual difference in employment between CES estimates and frame estimates (referred to as net birth/death residual) using an auto-regressive integrated moving average (ARIMA) time series model. This modeling produces a forecasted residual that is added on to the monthly employment estimates generated from the continuing units.

Concerns have arisen about whether this two-part modeling procedure adequately reflects cyclical variation in the birth of establishments and their employment growth across the business cycle. To a certain extent, the first component of the model will reflect cyclical variation because the growth rate estimated from continuing units will vary over the business cycle. The time-series model used in the second component of the procedure, however, does not include *any* covariates, whether they were related to changes over the business cycle or not. Consequently, the time-series component of the two-part procedure does not explicitly capture variation related to the business cycle. Research has been undertaken to determine if the time-series model could be improved by the inclusion of cyclically sensitive covariates, but to date, no variables that consistently improve the time-series model have been found.

4. Cost Containment of Data Collection

One of the most costly aspects of conducting the CPS and the CES is the data collection. To address concerns about rising costs, the CPS has implemented and explored various alternative modes of data collection. Likewise, the CES considered costs when designing its sample rotation scheme.

The CPS is administered by interviewers with the majority of the interviews being conducted over the phone from interviewers' homes. However, to reduce costs, approximately 15% of the phone interviews are conducted from centralized phone facilities. Among other criteria, cases are assigned to CATI facilities if it is difficult to hire interviewers in the specific geographic area.

Recently, the CPS has explored whether the survey could be conducted via the Internet. Collecting the CPS over the Internet would entail a complete redesign of the CPS questionnaire, particularly given that the CPS has several questions where respondents provide answers free form which interviewers then code into pre-specified categories. Consequently, as a preliminary step, a supplement was conducted in November 2013 requesting respondents provide an e-mail address at which they could be contacted. To evaluate the *feasibility* of a respondent completing the survey on-line, information also was collected in the supplement about how frequently individuals checked the address, whether the e-mail address was shared with others in the household (confidentiality concerns could preclude information collected in previous CPS interviews being shared with anyone other than the initial respondent), and whether respondents could be sent a link to the CPS survey. Finally, to access the *likelihood* of a respondent completing the CPS on-line, respondents also were asked on a scale from 1 to 10 how likely they would be to do the CPS on-line and if they had ever done a survey on-line. Unweighted analysis indicates that approximately 40% of households provided a legitimate e-mail address, although some of these addresses required cleaning and standardization. For only about 26% of households that provided an e-mail address did collecting CPS information via the Internet seem feasible (i.e. respondents indicated that they checked the provided address at least every two days, they did not share the address with others and it would be okay to send a link to the CPS survey). When information on respondents' own perception of how likely they would be to complete the CPS online was added, the proportion that might be expected to complete the CPS online was around 20%. Given these results, conducting the CPS survey on-line is being approached with caution. To reduce costs, the Internet might be used by interviewers to set up interview appointments and other forms of communication, however.

In the CES to reduce the cost and workload associated with enrolling new sample units, all units remain in the sample a minimum of 2 years. To ensure all units meet this minimum requirement, CES has established a "swapping in" procedure. The procedure allows units newly selected during the previous year but not reselected as part of the current probability sample to continue to remain in the sample. To account for these

units continued inclusion, newly sampled units within the same selection cell are removed. Approximately 60% of the CES private industry sample overlaps from the previous sample to the current sample.

5. Response Rate Challenges

In addition to the cost of collecting data, there also are concerns about response rates. For the CPS the concern is about declining response rates that are due either to the failure to find anyone at home at a housing unit or to the refusal of individuals to participate once they have been contacted. For the CES the focus has been on increasing the response rate and the timeliness with which responses are received.

Historically, the CPS has had a response rate over 90%. However, the average response rate in the first 3 months of 2015 was approximately 88%. To address concerns about the declining response rate, the Census Bureau has undertaken several measures to increase the likelihood that a respondent will be home when an interviewer attempts a contact. One measure includes providing interviewers' information from previous months indicating the day and time when unsuccessful attempts were undertaken and when an interview was actually completed. Using this information, it is hoped that interviewers will contact households during previously fruitful periods, increasing the likelihood of successfully reaching a respondent, reducing the amount of time spent trying to reach a respondent, and decreasing the number of potentially irritating phone messages interviewers leave on respondents' answering machines. To further enhance the probability of interviewers successfully contacting a household, the Census Bureau is conducting a "post card" experiment in which households are sent a post card with phone number for respondents to call to arrange a convenient interview time. Preliminary results of the test are quite promising, and the experiment is being expanded.

To increase the likelihood of actually completing a survey once a respondent has been reached, the Census Bureau has implemented several measures to encourage reluctant respondents to answer the survey. First, contact histories contains information on why respondents were hesitant or refused to respond in the past. Second, to help interviewers convert reluctant respondents, the Census Bureau has modified and intensified its refusal conversion training for all CPS interviewers. Finally, the Census Bureau has identified conversion specialists to whom they assign cases identified as potentially difficult to interview.

Several additional measures to increase CPS response rates are being investigated. The Census Bureau is accessing respondents' interest in localized data. If it is determined that respondents are interested, the Census Bureau will experiment with developing tailored handouts containing localized CPS estimates for interviewers to use. In another measure, BLS is investigating methods to identify which cases are most likely to be completed by centralized CATI, so cases can be assigned to optimize the response rate.

The CES was an early adopter of alternative modes of data collection, at least partially as a means to increase the response rate. Starting in 1991, CES began automated data collection by encouraging respondents to call a toll-free phone and entering their data in response to pre-record questions (Touchtone Data Entry). Currently, the CES has five main data collection methods: Computerized Assisted Telephone Interviewing (CATI), Electronic Data Interchange (where employers generate electronic files in a standard format directly from their payroll system and electronically transmitted the data to a collection center), Fax, Web, and Touchtone Data Entry. For the few establishments that do not provide data using one of these methods, data is collected either by mail, transcript, magnetic tape, or computer diskette. Centralizing data collection and allowing respondents a variety of ways to report has resulted in an increasing CES response rate over time.

How CES units are initialized into the survey, also has helped the CES increase its response rate. CES now has four regional Data Collection Centers (DCCs). These DCCs initially enroll units via the telephone and after enrollment collect the data for several months via CATI. After this initial period, where possible, data collection is switched to Touchtone Data Entry, Fax or Web. Very large firms receive special attention to ensure they respond, often being enrolled via personal visits. During the personal visits these large firms are assisted in establishing an Electronic Data Interchange for ongoing reporting. Finally, CES tries to maintain its response rate through an ongoing refusal conversion program conducted by the DCCs.

6. Maintaining Comparability Overtime – The Interface Between a Changing Society and the Surveys

A constant tension for the CPS and the CES is maintaining consistency in estimates across time, while also being attuned to societal changes. For the CPS, one area where concern arose was the use of the Internet by individuals looking for work. To be classified as unemployed in the CPS, an individual must be actively looking for work as indicated by having used at least one active job search method sometime in the prior four weeks. This information is obtained by interviewers asking respondents about all of the things they did to find work in the previous four weeks. Interviewers then recording these answers in pre-specified response categories that include both active and passive search methods.

In 1994, when the CPS last underwent a major redesign, relatively few individuals used the Internet to look for work. Consequently, the pre-specified CPS response categories for the job search methods question do not contain any mention of the Internet. Careful review of how the Internet is used to look for work by BLS staff determined that the Internet is a tool to look for work, and does not in of itself constitute a job search method. In order to determine whether individuals actively looked for work and thus should be classified as unemployed, interviewers need to determine what was done on-line. Review of how individuals used the Internet to look for work also revealed that the vast majority of Internet based activities could be coded into a pre-existing category. For instance, posting a resume to an electronic job board or completing an application on-line could be coded into the pre-existing category “sent out “resumes/filled out application”.

Monitoring of live CPS interviews along with focus groups of interviewers indicated that not all interviewers were consistently coding answers involving the Internet. To improve consistency, it was decided to expand interviewers’ instructions for the coding of job search methods. The expanded instructions were reviewed by BLS’s Technical Advisory Committee. They also were reviewed by a select group of interviewers in order to determine if the modified instructions were clear. The expanded instructions were implemented with the collection of May data in June 2015. The monitoring of interviews conducted from January 2015 through May 2015, indicate that while the expanded instructions should increase the precision of interviewers coding of responses, the labor force classification of very few individuals should be altered.

For the CES, given its emphasis on industry estimates, a tension arises when trying to accommodate new and emerging industries, while maintaining historical industry series. Starting with the release of May 2003 preliminary estimates, the CES converted to the 2002 North American Industry Classification System (2002 NAICS) from the 1987 Standard Industrial Classification System (SIC). Subsequently, with the release of the January 2008 preliminary estimates, the CES switched to 2007 NAICS. Since the release of the January 2012 preliminary estimates, the CES has used the 2012 NAICS. To avoid breaks in the time series adoption of the 2002 NAICS entailed, all national CES industry estimates were reconstructed back to at least 1990.

The reconstruction used BLS’s March 2001 Longitudinal Data Base of employers’ Unemployment Insurance records where units were assigned both a NAICS code and a SIC code. SIC to NAICS ratios for each SIC industry were constructed from this dual coded data. These ratios were used to map employment from the SIC series to their NAICS counterparts and the mapped employment estimates were then summed to obtain NAICS based employment estimates. Because BLS’s Longitudinal Data Base exists back to 1990, this methodology permitted NAICS based industry employment estimates to be constructed back to 1990. If a SIC and NAICS industry were found to be nearly identical, the NAICS based employment estimate was extended back beyond 1990 to when the SIC industry estimate was first started. BLS’s higher level sector, supersector, and two level goods-producing and service producing aggregations were extended back to 1939.

7. Expanding Demands for Labor Force Information

Another pressure on the CPS and CES is the ever increasing desire by users for more information about the functioning of the labor market and additional subgroups of the population. There are two approaches to address this demand. The first is to collect additional information. The second is to make available to the public more of the data that is already collected. The CPS and CES have utilized both approaches.

7.1 Expansion of Data Collection

The rapid rate at which the CPS data needs to be collected and tabulated, combined with the need to limit respondent burden and reduce the possibility of unintended context effects, precludes unlimited expansion of the CPS questionnaire. To balance the desire for additional information with data collection concerns, the

CPS asks one-time supplemental questions most calendar months of the year. The topics of these supplement often are directly related to the labor force experience of subgroups in the population such as veterans or individuals displaced from a job, however this is not always the case. To address changing labor markets and structure of work both the subject matter and specific content of supplements is reviewed.

Occasionally, a decision is made to add a few questions to the monthly CPS permanently. The most recent example of this is the addition of questions identify workers with certifications and licenses starting in January 2015. To accommodate the addition, of these questions, two underutilized educational attainment questions were removed from the monthly CPS. Prior to this addition, the next most recent addition to the monthly CPS was the addition in January 2009 of questions identifying whether individuals are disabled.

Finally in the wake of Hurricane Katrina, a “disaster block” has been embedded in the CPS instrument which allows for questions to be added quickly in the case of a natural or manmade disaster. This “disaster block” provides placeholders for questions with pre specified skip instructions. Having pre-existing skips hopefully mitigates the probability of introducing unintended error into the collection of the regular monthly data.

The CES has met the demand for additional data by expanding the group of employees for which it collected hours and earnings data. For most of its history, the CES only collected hours and earnings data for production and non-supervisory employees. Concurrent with the release of the January 2010 data, CES began releasing hours and earnings data for all employees.

7.2 Expansion of Data Released

In response to the request to make existing data more available, the CPS has increased the amount of information in the monthly Employment Situation press release, increased the number of news releases on groups of special interest, and generated some additional special topical reports. For instance, the monthly Employment Situation now includes tables on the labor force status of veterans, the disabled, and individuals born outside of the U.S. Tables within the monthly Employment Situation press release were expanded to include seasonally adjusted estimates for Asians, adult Hispanic men, adult Hispanic women, and women age 55 and older (previously only not-seasonally adjusted estimates were available for these groups).

In response to users, in 2015 CES began to release in a single table *all* of the estimates generated for a month. Although the first preliminary CES estimates typically garner the most public attention, the first monthly estimates are revised a number of times before they are finalized. These revisions are due to the ongoing receipt of data in the two months subsequent to the release of initial estimate, annual benchmarking of the monthly estimates to universe employment counts, and the seasonal adjustment process. Prior to 2015, the only way for users to obtain previously published estimates was to assemble the estimates themselves from old press releases. Provision of all the estimates for a month in a single table allows users to more readily compare estimates generated in a similar manner and provides a more complete historical picture.

Finally, both the CPS and the CES recently increased the accessibility and visibility of their data. Both programs have increased the number of tables users can access online and increased the number of individual estimates users can download via the BLS online data base. Online tables have been modified to insure they are compatible with software used by the visually impaired. Starting in June 2012, BLS began providing information via Twitter. As of 2014, the BLS has had an Application Programming Interface that allows users to write apps that insert published BLS statistics into their graphics or other products. Furthermore, starting in the fall of 2015, BLS plans to add a fully-updated chart package to accompany the Employment Situation

8. Conclusion

While issues and opportunities will change over time and vary by items measured, statistical programs always face a need to balance the trade-offs between continuity and relevance in their decisions. In that sense, the challenges we face today for the estimates presented in the Employment Situation are no different from those faced by our BLS predecessors and by the heads of other statistical programs. We hope this discussion proves useful to others. We welcome the input of our users and our statistical community colleagues on how best to meet these challenges, as we strive to provide our users with gold-standard data to help them make the best decisions for their businesses, communities, families or constituents.

