ENSURING DATA QUALITY IN THE JOB OPENINGS AND LABOR TURNOVER SURVEY

Charlotte Mueller, Matthew Balazik, Bureau of Labor Statistics
Charlotte Mueller, BLS, Suite 4840, 2 Massachusetts Ave., NE, Washington, DC 20212

Key Words: Turnover; Labor Market; Labor Demand; Hires and Separations; Job Openings; Outlier.

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

1. INTRODUCTION

The Bureau of Labor Statistics (BLS) has begun a new survey to collect and publish job openings and labor turnover data. This new survey, the Job Openings and Labor Turnover Survey, or JOLTS, began collecting data in March 2000 and will soon publish estimates for the first time. Published estimates will include rates and levels for job openings, hires, and separations. These data will add to the United States employment picture drawn by two other BLS products: the employment level and the unemployment rate.[1]

A high priority for JOLTS is data quality. With a sample size of only 16,000 establishments (small by BLS standards), high response rates and data quality are very important. Beginning with the first month of collected data (March 2000), the JOLTS staff has edited and reviewed the data. Being a new survey has allowed great flexibility and opportunity to choose how to edit and review the data, and how to handle the output from these processes. Methods have been devised, implemented, and revised throughout the first year of data collection.

2. DATA COLLECTION

Each month, the data are collected for approximately two weeks in the data collection center (DCC) in Atlanta, Georgia. The JOLTS DCC includes 22 contract interviewers, 2 contract supervisors, and a BLS manager and assistant manager. Modes of collection for JOLTS are as with many surveys: computer assisted telephone interviewing (CATI), touchtone data entry (TDE), fax, and mail. A small number of units are collected through special arrangements. The goal is to have 75 percent of the responding units report via TDE, although reporting by fax seems to be gaining interest with respondents. The majority of the remaining units would report via CATI. Each month at the end of the collection period, a snapshot of the database is copied to the BLS national office in Washington, DC, where the JOLTS staff begins the steps to publication.

3. OUTLIER DETECTION

The first step after receiving the monthly data file is to screen and adjust for outliers. This important first step in ensuring data quality (and hence the resulting estimates) occurs before point estimation. Although this step is not yet implemented in JOLTS, it soon will be. For the time being, the outliers are identified and handled during the editing process. For each data element, the ratio of the reported value to the reported employment is screened for outliers, as well as the month-to-month change in those ratios. For employment, the ratio of current to previous employment is screened. Initially, the values will be compared within a sampling stratum or collapsed sampling stratum. (In JOLTS, the sampling stratum is ownership/region/industry division/size class.) An alternative to identifying outliers within strata is to develop parameters for outlier detection using multiple months' data.

Two outlier detection methods currently under consideration are Rosner's Test and Grubb's Test. After detection, the outliers likely will be adjusted using Winsorization. Of course, the outlier data will be investigated first to ensure that the data values are not reporting or keying errors.

3.1 Rosner's Test

Rosner's generalized Extreme Studentized Deviate (ESD) many-outlier procedure (Rosner, 1983) is used to detect k upper or lower outliers from n data points. The null hypothesis is that there are no outliers in the dataset. There is one alternative hypothesis each for 1, 2, …, k outliers in the dataset. To begin, the mean and standard deviation are calculated using the full n observations. Then, the

[1] A number of papers about JOLTS were presented at the 2000 Joint Statistical Meetings in Indianapolis. See Clark and Hyson, Crankshaw and Stamas, Mueller and Phillips, and Mueller and Wohlford for more information about JOLTS.

observations are ordered from 1 to n based on their deviation from the mean, with x_n being the point furthest from the mean. The test statistic, R_1, is then computed. That is, $R_1 = (\max|x_i - \bar{x}|)/s$, where $\bar{x}$ is the usual sample mean, and s is the usual sample standard deviation. The critical value λ_{j+1} is calculated for each possible number of outliers j = 0, 1, …, k-1. The probability evaluation is Pr $[(R_{j+1} \leq \lambda_{j+1}) \mid H_j]$, j= 1, …, k-1. If R_1 is greater than the critical value, then x_i is an outlier. At that point, x_i is deleted from the dataset, and the mean, standard error, and test statistic are recomputed on the remaining n-1 data points, using the alternative hypothesis of two outliers. This process is repeated until the test fails to reject the null hypothesis. All data points tested up until the point of failing to reject the null hypothesis are considered to be outliers. That is, if n-4 is the first ranked data point to fail to reject, then n, n-1, n-2, and n-3 are considered outliers. For the full paper, see Rosner (1983).

3.2 Grubb's Test

Grubb's Test is another iterative order statistic approach to detect the presence of outliers. The natural log transformation is applied to all data points, then the resulting values ranked in ascending order. The usual mean and standard deviation for the full dataset are calculated and used to compute the upper and lower Tau Statistics, $T_{lower} = (\bar{x} - x_1)/s$ and $T_{upper} = (x_n - \bar{x})/s$. The critical Tau value is based on the sample size and selected alpha level. The null hypothesis is again that there are no outliers, and the alternative hypothesis is that the data point being tested is an outlier. (See NIST references.)

3.3 Winsorization

If any outliers are detected in the JOLTS dataset, they must be managed within their stratum or collapsed stratum. In the Winsorization method, the n observations are again ordered from smallest to largest. Then the k-1 most extreme data points are replaced by the kth data point, where x_k is the first non-outlier value. Within one dataset, lower or upper values, or both, may be Winsorized. The (lower) Winsorization of the dataset $\{x_1, x_2, \ldots, x_{k-1}, x_k, \ldots, x_{n-1}, x_n\}$ is $\{wx_1, wx_2, \ldots, wx_{k-1}, x_k, \ldots, x_{n-1}, x_n\} = \{x_k, x_k, \ldots, x_k, \ldots, x_{n-1}, x_n\}$. For more details on Winsorization, see Cox (1995).

4. EDITS AND PARAMETERS

To ensure data with the fewest possible errors, additional edits are needed after outlier adjustment. Not all errors can be found, of course, but the largest changes and differences can be output for verification or correction. During point estimation, the first edit compares each sample unit's reported employment to the unit's frame employment. These changes are especially important to identify so that nonresponse adjustment through application of nonresponse adjustment factors (NRAFs) does not overly inflate one unit's influence on the estimates. In addition to editing reported employment, the level for each data element (job openings, hires, quits, layoffs & discharges, other separations, total separations) is compared to the reported employment for the unit. Different parameters for the edits were derived using previously collected data from a JOLTS pilot study. The parameters vary based on size class and data element. Once enough months of data have been collected, the reported employment also will be compared to the sample unit's year-ago employment in order to accommodate seasonal effects.

5. MICRODATA REVIEW

The purpose of microdata review is not necessarily to challenge the responses collected from contacts at establishments. Rather, it is to investigate reported data that appear out of the ordinary. An interviewer might recontact the establishment to probe for an explanation from the respondent, and in some instances revise their reported data if needed. In some cases, interviewers simply may have made a keying error during the interview. For TDE-reporting establishments, the respondent may have made an error when entering the data on the phone.

One advantage of a relatively small sample is the ability to intensely review the microdata. One year into data collection, only about half of the 16,000-establishment sample was initialized into the database. Working with 9,000 establishments, we are able to study the microdata and evaluate the collectibility and quality of the JOLTS data. Once the monthly end-of-collection snapshot is taken and the data file is transferred to the national office, analysts review the data. For at least the first year, all records with reported data were reviewed by industry. Each analyst is the primary reviewer of at least one industry every month and each industry is reviewed by two different analysts. All data inconsistencies, irregularities, or atypical activity for the industry are investigated.

In addition to the employment edits performed during point estimation (see Section 4 above), the analysts edit all the other data elements (job openings, hires, and separations). The month-to-month change in the level for each data element is screened for unusually large changes. Parameters for edit failures are based on the size class of the establishment and the data element under scrutiny. The change in level for each data element is compared to the establishment's reported employment for the month. In order to fail the edit, there must be a large change in the level that is greater than a pre-set percentage of the month's reported employment. All of these edits are performed by size class and data element. Each establishment failing one or more edits is output for review; flags indicate which data element(s) failed the edit(s). The outputs from these edits are distributed to the analysts for review according to the month's industry assignments. The analysts look up each failed unit in the interviewers' data collection system[2], where they study the establishment's data history and read all the notes entered by the interviewer for the unit in question.

Beyond the edits designed to catch large fluctuations in reported data, the JOLTS staff runs programs to assist analysts by identifying establishments with inconsistent responses. For example, when a respondent's reported job openings equals their reported hires for several months in a row, it may mean that the respondent does not understand the reference periods of the data elements. Because many establishments manage to fill job openings within the same month the openings occur, the number of hires for a month can reasonably equal the cumulative number of job openings throughout the month. However, the job openings data element in the JOLTS survey includes only openings that remained unfilled on the last business day of the month. So, in the case above, job openings should equal zero if all the openings were filled during the course of the month. For most establishments, the number of job openings will be less than the number of hires for the month. Without this check, the job openings count would be biased upward.

Other suspect cases that are automatically flagged for investigation include large establishments (employment over 1,000) that have reported a zero for one or more of the data elements. The likelihood of such an occurrence for large establishments is fairly low and warrants further investigation. Likewise, small firms (employment under 15) that have reported "data not available" for any elements, are flagged for investigation. It should be relatively easy for respondents in small firms to report the requested JOLTS data elements. Before sending the record back to the DCC for investigation, analysts review the interviewer's notes to look for an explanation of why any data element was not reported.

After reviewing the output from the edits, analysts review the industry listing of the month's microdata (mentioned above). Still looking for anomalies, analysts check that large reported turnover matches the industry trend or that an explanation is provided in the interviewer's notes. Analysts also look for reported values that seem unusually large, for zeroes reported where "data not available" seems more appropriate, and for establishments with many hires or separations in recent months, but without corresponding changes in reported employment.

Again, this thorough review is possible only with a small sample size, and only with part of the sample active. As the active sample increases, we will rely less on manual review and more on the automated edits. This progression is not inherently bad, given the trend in the past: increasing automation has improved the consistency of how data are reviewed. As a matter of fact, the lists of problem records generated by hand and by computer have high overlap. But, the automated filters tend to report more cases because of their ability to easily compare data *between* months as well as *within* the current month.

Additional edit programs screen the microdata for errors that need no investigation, but must be fixed. This includes: 1.) Records in which the sum of the separations (quits, layoffs and discharges, other separations) does not equal the total separations on the file; 2.) Records for which usable data is present but the record is not coded with a usable status code; and 3.) Records that are coded as usable but lack data. All of these types of errors are forwarded to the DCC for correction.

Microdata review is complete once the national office staff records all errors and unresolved items into an electronic spreadsheet shuttle form and sends it to the DCC. There, interviewers have three working days to review the cases in question. If necessary, they may contact the establishment and discuss the reported value(s) that failed the edit and review process. The interviewer may have to re-

[2] The data collection system uses the Blaise software package written by Statistics Netherlands. Extensive survey-specific modifications have been made to the software in order to support the JOLTS program.

explain definitions or reference periods and confirm or correct the data. Then the interviewer must document the change or the reason for the irregular data by entering a note in the system, or by selecting an appropriate comment code. The notes and comment codes let the national office know that the reported data are verified and are not the result of a keying or reporting error. The DCC then enters into the shuttle form the action taken for each listed establishment, and returns the form to the national office.

During the monthly data review process, the JOLTS staff looks for common problems. Anything common across several establishments or interviewers is noted on the DCC shuttle form. Larger scale data collection problems are addressed through additional training for the interviewers. This training includes directions to increase consistency among interviewers in using the data collection system, as well as training geared toward potential respondent misunderstandings about JOLTS data. The training is either forwarded to the DCC for the managers to conduct, or conducted by the national office staff during the next visit to the DCC. (The national office tries to visit the DCC on a monthly basis.)

Another possible outcome of the microdata edit and review process is the identification of necessary changes to the data collection systems (CATI and TDE) or the data collection form. If the same type of problem occurs regularly or is widespread across interviewers, the national office may request that a change be made to the system to eliminate that type of error. One example is when we adjusted the CATI system to automatically calculate and enter the total separations once the three individual types of separations were entered. This eliminates arithmetic errors and reduces the number of tasks the interviewers have to perform. Although not a part of microdata review per se, continual examination of the data collection forms is necessary in order to evaluate their effectiveness.

6. JOLTS IN THE NEAR FUTURE

As with any good program, all aspects of JOLTS will be monitored, and improvements made when needed. Nonsampling errors due to system problems, training needs, keying errors, or respondent misunderstandings will be addressed. To do this, the JOLTS staff will continue to edit and review data as thoroughly as possible. The increased reliance on automated editing and checking of data will help maintain the quality of the data even as the number of active sample units increases. As we move toward publication in early 2002, extra care will be used when reviewing the microdata and analyzing the estimates. As more data are gathered, the edit parameters and types of edits will be refined to flag all values needing review, but not too much more, so as to thoroughly analyze the data without unnecessarily increasing the workload.

During the series' experimental two years, the data collection systems (CATI and TDE) will be monitored continually, and the data collection forms and the interviewers' CATI scripts should be re-evaluated. Because JOLTS is an experimental series for the first two years, the systems, forms, and scripts should be flexible. One way to fully evaluate the data collection form and the interviewer scripts would be to conduct a formal Response Analysis Survey (RAS) in which JOLTS staff visit selected respondents, and ask in detail what they reported and what definitions they used. The main focus is to determine if the respondent understands the definitions and reference periods, and to assess the respondent's willingness and ability to provide the data.

As during the first two years of the JOLTS survey, staff will continue to answer email requests to the JoltsInfo@bls.gov email group, and phone calls to the JOLTS help line (202/691-5870). The JOLTS page on the BLS website (www.bls.gov/jlthome.htm) will be updated, and estimates will be available on-line once publication has begun.

7. REFERENCES

Clark, Kelly and Rosemary Hyson. "Measuring the Demand for Labor in the United States: The Job Openings and Labor Turnover Survey." ASA Papers and Proceedings, 2000.

Cox, B., D. Binder, B. Chinnappa, A. Christianson, M. Colledge, and P. Kott (1995), "Business Survey Methods," Wiley, pp. 510-512.

Crankshaw, Mark and George Stamas. "Sample Design in the Job Openings and Labor Turnover Survey." ASA Papers and Proceedings, 2000.

Grubbs, Frank. "Sample Criteria for Testing Outlying Observations", Annals of Mathematical Statistics, Vol. 21, Issue 1, Mar., 1950: 27-58.

Mueller, Charlotte and Mary Anne Phillips. "The Genesis of an Establishment Survey: Research and Development for the Job Openings and Labor Turnover Survey at the Bureau of Labor Statistics." ASA Papers and Proceedings, 2000.

Mueller, Charlotte and John Wohlford. "Developing a New Business Survey: The Job Openings and Labor Turnover Survey at the Bureau of Labor Statistics." ASA Papers and Proceedings, 2000.

National Institute of Standards and Technology (NIST), Information Technology Laboratory, Statistical Engineering Division, Dataplot, Vol. 1, Auxiliary Chapter, Grubbs Test. 18 May 2001. <www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/grubtest.htm>

National Institute of Standards and Technology (NIST), Information Technology Laboratory, Statistical Engineering Division, Engineering Statistics Handbook, Chapter 1, Section 3, Subsection 5.17, Grubbs Test for Outliers. 18 May 2001. <www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm>.

Rosner, Bernard (1983), "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, Vol. 25, No. 2, May 1983.

Snedecor, George W. and Cochran, William G. (1989), *Statistical Methods, Eighth Edition*, Iowa State University Press.