SAMPLE DESIGN IN THE JOB OPENINGS AND LABOR TURNOVER SURVEY

Mark Crankshaw and George Stamas* U.S. Bureau of Labor Statistics
Mark Crankshaw, Suite 4985, 2 Massachusetts Avenue NE, Washington, DC 20212

Key Words: Sample Design, Establishment Survey

1. INTRODUCTION

The Job Openings and Labor Turnover Survey (JOLTS) is a new program being fielded by the Bureau of Labor Statistics to compile measures of job openings and employee turnover for nonfarm businesses using a sample of 16,000 establishments. The job opening, employee turnover, and employment data are collected on a monthly basis. Employee turnover includes hires, and three categories of separations—quits, layoffs and discharges, and other separations. Most sample members will rotate out of the sample after participating for 18 months. Estimates will be produced for broad industry groups and Census regions.

In this paper we will discuss JOLTS sample design, JOLTS estimators, sample coordination with other BLS surveys using permanent random numbers, and issues arising from sample selection using the Standard Industrial Classification (SIC) system as a strata identifier and estimating for North American Industry Classification System (NAICS) sectors.

JOLTS is an establishment survey designed to assess rates of job openings and labor turnover relative to employment for key industrial sectors of the United States. Abraham (1983) discusses the job-openings rate as a labor demand companion statistic to the widely known labor supply measure, the unemployment rate. Clark, Cohen and Hyson (2000) discuss the concepts behind the measures produced by JOLTS in relationship to recent literature. Mueller and Phillips (2000) discuss the origins of JOLTS, and Mueller and Wohlford (2000) discuss JOLTS operations. Goldenberg and Phillips (2000) describe the extensive effort that went into defining concepts and developing the survey instruments for JOLTS. The scope of JOLTS is the same as that of the BLS monthly Current Employment Statistics (CES) survey--payroll employment in nonagricultural industries in the private sector, Federal and State and local government, across the 50 states and the District of Columbia. JOLTS will ratio adjust or "benchmark" estimates to employment estimates from the CES. (Butani, et al., 1997)

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

2. SAMPLE DESIGN

JOLTS uses the same frame as the CES, the BLS Longitudinal Data Base (LDB) as modified by CES. The LDB, is a data base of U.S. business establishments covered by state or federal unemployment insurance (UI) laws. The file includes a record for every active UI account issued by any state. It also includes a record for individual reporting units operating under that UI account, or an indication that the account has multiple work sites but that the employer is not required, by virtue of employment size, or refuses to provide individual work site data to the state. As the name suggests, LDB records are linked longitudinally. CES assigns a permanent random number to each record on the frame. While CES samples UI Accounts from this frame, JOLTS selects a sample from the reporting units—either the UI account or in cases where UI accounts have multiple reporting units, those subunits. These reporting units correspond closely to individual physical locations or work sites. The JOLTS frame excludes records for private household workers (SIC 8811), and those from agriculture, forestry and fisheries, with the exception of agricultural services (SICs 074, 075, and 078). Any records without employment in any of the six most recent months on the frame are also excluded. We selected the current JOLTS sample from a frame that had been updated for data available through March 1999.

The JOLTS sample was stratified into groups defined by the four Census Regions, six size classes based on establishment employment, and eleven industry divisions defined by ownership and Standard Industrial Classification. (See Table 1.) Large establishments were sampled with virtual certainty, nearly 100 percent.

We allocated the JOLTS sample using the Neyman method. Variance information on employment by strata was available from the survey frame and other establishment surveys, however, similar information was not available for the key survey items job openings, hires and separations. Using preliminary JOLTS data from a small pilot survey conducted for BLS by the Westat Corporation, we saw that the mean employment within strata approximated within strata variance of employment and was proportional to variance of other key items. (Levin, et al., 2000) Following text book suggestions, we used the mean employment within a stratum for allocation. (Cochran, 1977) As a result of

this allocation, and as is the case in most establishment surveys, larger establishments were selected with greater probability than were smaller establishments. When the sample is updated in a year, we might be able to allocate more efficiently because we will have the additional information from months of survey operation.

Table 1. Industry Composition of JOLTS Strata

Division		Ownership and SIC
Private	**ID**	**(Own=5)**
Mining	10	1000-1499
Construction	20	1500-1799
Durable Goods Mfg	31	2400-2599, 3200-3999
Nondurable Goods Mfg	32	2000-2399, 2600-3199
Transport & Utilities	40	4000-4999
Wholesale Trade	50	5000-5199
Retail Trade	60	5200-5999
Finance,Ins&RealEstate	70	6000-6799
Services	80	7000-8999, 074, 075, 078 except 8811
Government		**(Own=1,2, or 3)**
Federal	91	Federal (all sics)
State & Local	92	State, Local (all sics)

To limit burden on any one respondent to the extent possible, with some consideration for cost of initiating an employer to the survey, we chose a rotating panel design. The design has 18 noncertainty panels. Each unit is in the sample for 18 months, with a new panel introduced each month. After 18 months in the sample, a panel will be replaced with a new panel. We selected a minimum of one unit per stratum per panel. All sample sizes were rounded to a multiple of 18 to allow for 18 equally sized panels. The design also includes a virtual certainty panel of sample units that remain in the sample. Units are selected to this panel by virtue of past employment size, 100 percent sampling rates indicated by the allocation or because the sampling rate is so large as to make rotating impractical, or because of minimum sample requirements. With resources available for an approximate sample size of 16,000 units, we selected the sample and divided it across 18 noncertainty panels of equal size, plus a virtual certainty panel containing all units selected with near certainty.

To facilitate JOLTS sample rotation and survey coordination with the Occupational Employment Survey (OES), the CES, and future JOLTS samples, sample selection was done using Permanent Random Numbers (PRN). Each establishment record on the frame has been randomly assigned a PRN. The OES and CES each have a predetermined starting point based on the assigned PRN, at which they begin to select units for their sample. The JOLTS sample was selected to limit overlap between surveys by choosing a PRN start value after those for the CES and the OES samples. The sample frame was sorted by PRN within strata. The first n_h units following the starting point within the h^{th} stratum were selected, where n_h is the sample size for the h^{th} stratum. Each stratum in each panel, p, has a sample size $n_{ph}=n_h/18$. The units with the first n_{ph} JOLTS PRNs go into stratum h for panel one, the second n_{ph} PRNs go into stratum h for panel two, and so on until the 18^{th} panel has been populated for stratum h. Subsequent JOLTS samples will use a starting point directly following the ending point within each stratum of the current JOLTS sample.

3. ADJUSTMENTS FOR NONRESPONSE

JOLTS will produce two sets of estimates each month—a current month preliminary estimate and a prior month revised estimate that reflects additional sample responses and possibly corrections to responses used in the preliminary estimate. We refer to these as first and second closing estimates. At the time that we produce first closing (preliminary) estimates for the current month we will also produce second closing (revised) estimates for the previous month.

Even with a second closing, the survey will have nonresponse issues to address: failure to enroll into the survey; nonresponse during monthly collection; and item nonresponse. The first two types of nonresponse will be compensated for with nonresponse adjustment factors (NRAFs); the third will be addressed through a hot-deck imputation procedure that uses a nearest neighbor approach. Rather than borrow the missing item from the nearest neighbor, JOLTS imputation will borrow a ratio—for example, Job Openings/Employment, Hires/Employment, etc.—and multiply that borrowed ratio by the employment value of the recipient record. All of these adjustments will be carried out at the stratum or collapsed stratum level.

3.1 Establishment Nonresponse

NRAFs will be calculated every month based on the total sample size and the number of respondents available that month. They will be calculated across noncertainty panels and separately for the certainty panel. These factors are calculated by summing selection weights of viable sample cases and, separately, the selection weights of the usable sample cases.

$$NRAF_{ch,p\in(1,18)} = \frac{\sum_{i\in ch,viable,p\in(1,18)} w_{i,ch}}{\sum_{i\in ch,usable,p\in(1,18)} w_{i,ch}}$$

$$NRAF_{ch,p=0} = \frac{\sum_{i \in ch, viable, p=0} w_{i,ch}}{\sum_{i \in ch, usable, p=0} w_{i,ch}}$$

where w is the selection weight, i is the individual establishment, ch is the collapsed stratum, *viable* are all units in the sample excluding the out-of-scope, *usable* are all units responding with valid data, and p is the panel 0-18, panel *0* being the certainty panel. An alternative method, in place for the early months of data collection, combines sample responses from all panels fielded, adjusts the selection weights from the noncertainty panels to reflect the number fielded, and then calculates the NRAFs across responses from all panels.

3.2 Item nonresponse

JOLTS will impute values for Job Openings, Hires, Quits, Layoffs and Discharges, and Other Separations wherever these fields are missing data in the responding sample units, using a nearest neighbor approach. Responses are pooled *across* panels and sorted by strata, and by employment within strata. Records with missing values for a given item are designated as recipients for that item. Records with responses for the item serve as the donor pool. The donor selected for any given recipient will be the record with a response on the item and the smallest difference in employment from among the donor population. If e_i is employment of the recipient record, then we select as a donor the record j' from the donor pool, D, such that $|e_i - e_{j' \in D}| < |e_i - e_{j \in D}|$ where $j' \neq j$. Once a recipient record has been paired with its donor, the procedure imputes a value by calculating the item ratio (item response/employment) from the donor and multiplying that ratio by the employment from the recipient record.

4. ESTIMATORS

4.1 Method

JOLTS estimates for Job Openings, Hires and three types of separations will be produced using ratio estimation. This method increases precision in an estimate of Y by taking advantage of the correlation between y_i and an auxiliary variate, x_i. (Cochran, 1977) For JOLTS, the auxiliary variate (x_i) is employment, which is correlated with the other items (y_i), collected from the JOLTS sample. The population total X of the x_i should be known. We use CES employment estimates which, while they are estimates, are timely and comprehensive.

Define:

p = panel = {1,2,...,18,0}

0 = certainty panel

t = current month

d = IndustryDivision

r = CensusRegion

est_cell = estimating cell some combination of region and division

$BMKF$ = Benchmark Factor

CES = U.S. Industry Division Employment

w_i = weight reflecting all adjustments from a sample unit, i

e_i = Reported employment from a sample unit, i

$\hat{E}$ = Estimated employment

JO_i = Reported job openings from a sample unit, i

$\hat{JO}$ = Estimated job openings

x_i = Reported value for item X from a sample unit, i where X is Employment, JO, Hires, Quits, L & D, Other Separations or Separations (total)

$\hat{X}$ = Estimated X

Each noncertainty panel, when combined with the virtual certainty panel, is representative of the universe. We make the following calculations for each combination of the certainty panel plus one noncertainty panel. After doing the calculations for each noncertainty panel, we obtain the overall item estimate by summing the individual panel item estimates and dividing by the number of noncertainty panels.

To produce estimates, we calculate for each industry division a weight adjustment, or benchmark factor, *BMKF*. This equals the CES industry division estimated employment for the U.S. divided by the sum of the weighted sample employment for the division.

$$BMKF_{i \in d} = \frac{CES_d}{\sum_{i \in d} w_i * e_i}$$

The benchmark factor ratio adjusts a simple expansion estimator of employment to CES employment estimates at the Industry division level (as defined for JOLTS). The employment estimator can then be expressed as:

$$\hat{E}_d = \sum_{i \in d} BMKF_i * w_i * e_i = CES_d$$

The estimator for employment for any estimating cell is:

$$\hat{E}_{est\_cell} = \sum_{i \in est\_cell} BMKF_i * w_i * e_i$$

The estimator for Job Openings can then be viewed as a ratio estimator. As can be seen after the second equal sign in the formula below, it uses the benchmark factor to ratio adjust a simple expansion estimator.

$$\hat{JO}_{est\_cel} = \frac{\sum_{i \in est\_cel} w_i * JO_i}{\sum_{i \in est\_cel} w_i * e_i} * \hat{E}_{est\_cel} = \sum_{i \in est\_cel} BMKF_i * w_i * JO_i$$

We can generalize to the other data items: Hires, Quits, Layoffs and Discharges, and Other Separations. Each item can be represented by x in the following equation.

$$\hat{X}_{est\_cel} = \frac{\sum_{i \in est\_cel} w_i * x_i}{\sum_{i \in est\_cel} w_i * e_i} * \hat{E}_{est\_cel} = \sum_{i \in est\_cel} BMKF_i * w_i * x_i$$

The overall current monthly estimator ($\hat{X}_{t,est\_cell}$) for any item x is the mean of the monthly panel estimates. For example, the overall estimate of Job Openings, x, is the sum across panels of the estimated number of job openings in the estimating cell ($\sum_{p=1}^{\#p} \hat{X}_{t,p,est\_cell}$), divided by the number of noncertainty panels (#p).

$$\hat{X}_{t,est\_cell} = \frac{\sum_{p=1}^{\#p} \hat{X}_{t,p,est\_cell}}{\#p}$$

where #p is the number of panels (not including certainty) available for the estimates (maximum of 18). Note that at the industry division level:

$$\hat{E}_{t,est\_cell \vDash d} = \frac{\sum_{p=1}^{\#p} \hat{E}_{t,p,est\_cell}}{\#p} = \frac{\#p * CES_{t,est\_cell}}{\#p} = CES_{t,est\_cell \vDash d}$$

An alternative method, in place for the early months of data collection, pools sample responses across all panels fielded, and then applies the estimator. This produces one set of estimates rather than a set for each panel, so there is no need to average.

4.2 Rates and over-the-month changes

Estimators for Job Openings, Hires and the various separations rates are simply ratios of the estimators for each respective item divided by employment applied at any given estimating cell level.[1] Estimates of over-the-month change will be calculated as the estimated rate for the current month minus the rate for the previous month. Each month we will update estimates for the previous month to reflect any additional sample responses. These are known as second-closing estimates. They will be published as monthly revisions, and they are the estimates that will be used to calculate over-the-month change.

4.3 Variance estimation

We intend to use a sample replication technique to estimate the variance of JOLTS estimators. One of the simplest to implement is the Jackknife. Because we expect item nonresponse, and intend to adjust for that using nearest neighbor imputation, we are exploring the use of the Partially-Adjusted Jackknife described by Chen and Shao (1999). Their research has demonstrated that with their test data this technique improved on the understatement of variance resulting from ignoring that some data were imputed, and the overstatement that resulted from re-imputing within each jackknife. Variances will be computed using this technique for each estimating cell and each published cell for the estimated level and rate of each JOLTS item.

5. ACROSS SURVEY SAMPLE COORDINATION

We wanted to minimize burden of any one employer across BLS surveys to the extent possible, while trying to minimize overall burden, through an efficient survey design. To this end, we coordinate JOLTS sample selection with that of CES and the Occupational Employment Statistics (OES) program through the use of PRNs. Several design considerations limit our ability to do this. In the case of CES, the sampling unit is the UI account while JOLTS uses the reporting unit. Further confusing the issue is that the lower bound for certainty selection in CES is considerably lower than that for JOLTS. Still, we can use PRNs to coordinate the selection of UI accounts that have only one

[1] Research is under way in defining and calculating rates. By definition, job openings are not accounted for in employment counts. We will consider estimating the Job Openings rate as Job Openings divided by the sum of Employment and Job Openings as well as other variations that will attempt to mitigate the effect of different reference periods for different items.

Table 2. JOLTS and CES Sample Match Comparison by Industry Division (ID).

ID	CES Certainty	Grand-fathered Singles	Other Single Matches	No Match Singles	Singles	Multis	Jolts Sample units
10	13	11	40	171	235	87	322
20	38	32	29	700	799	38	837
31	115	159	27	746	1047	385	1432
32	65	81	17	569	732	279	1011
40	62	21	3	427	513	305	818
50	26	28	3	748	805	112	917
60	46	30	9	1796	1881	925	2806
70	58	18	5	565	646	307	953
80	497	201	119	2755	3572	845	4417
91	167	0	0	306	473	0	473
92	795	0	0	1260	2055	0	2055
	1882	581	252	10043	12758	3283	16041

reporting unit on the frame. The CES sample also includes cases that have participated in the CES but were not selected for the redesigned sample. These cases were retained or "grandfathered" into the CES sample if they had 100 or more employees over the previous 12 months. JOLTS could not be coordinated around these cases using the PRNs. We have matched against the CES sample the JOLTS sample units that are from single-unit UI accounts (80 percent). Table 2 presents the results our efforts at coordinating this first JOLTS sample and the current CES sample. The overwhelming majority of the single-site employers found in JOLTS were not in the CES sample (78.7 percent). Of the 21 percent that were in the CES sample, 15 percentage points of that were selected for CES with certainty, less than 5 percentage points were grandfathered in, and only 2 points matched for other reasons, including JOLTS allocation requiring a large sampling fraction.

Table 3. JOLTS and OES Sample Match Comparison by Industry Division (ID) in the Private Sector.

ID	OES Certainty	Other Matches	No Match	Jolts Sample units
10	36	34	252	322
20	45	41	751	837
31	316	123	993	1,432
32	165	94	752	1,011
40	152	91	575	818
50	71	66	780	917
60	106	294	2,406	2,806
70	161	98	694	953
80	763	45	3,249	4417
	1,815	1,246	10,452	13,513

Sample overlap between JOLTS and OES is easier to analyze. For private sector employers, OES samples reporting units, as does JOLTS. OES has a sample of 1.2 million establishments over a three-year survey cycle. As would be expected, given the very large sample size, the lower bound for certainty cases is considerably lower in OES than in JOLTS. Table 3 presents the results of sample coordination between JOLTS and the OES 2000 sample. The table shows that 23 percent of JOLTS private sector sample units are also in the OES. However, when OES certainty units are excluded only 11 percent of JOLTS sample units are also in OES.

6. ESTIMATING ON NAICS

Along with other BLS economic statistics as well as those at other federal agencies, the JOLTS survey will eventually publish estimates on NAICS. Mikkelson, Morisi and Stamas (2000) describe the NAICS conversion process for the UI Covered Employment and Wages Program, the program that provides the LDB file which serves as the JOLTS sampling frame. Because the LDB was not fully coded with NAICS when the initial JOLTS sample was selected, stratification and the resulting sample allocation and selection were defined by SIC.

In order to produce estimates on NAICS, JOLTS must deal with two issues. First, sample units must have NAICS codes assigned. Each record on the LDB and in the sample has a unique number assigned—an LDB number. To assign NAICS codes, sample records can be linked back to the LDB by LDB number once the file is updated to include NAICS. The second issue deals with the CES as a benchmark source for JOLTS.

The CES will convert to NAICS after JOLTS is scheduled to release data based on NAICS. Until CES converts to NAICS, and a JOLTS sample is selected, based on combined NAICS sector stratification, benchmark factors will be calculated as described above in section 4. Until CES adopts NAICS for publication, JOLTS will continue calculating the benchmark factors on an SIC division basis and those factors will be assigned to each sample unit by SIC division. For any item X, the estimator will be:

$$\hat{X}_{NAICS} = \frac{\sum_{i \in NAICS} w_i * x_i}{\sum_{i \in NAICS} w_i * e_i} * \hat{E}_{NAICS} = \sum_{i \in NAICS} BMKF_i * w_i * x_i$$

where the summation is across any sample unit, *i*, in NAICS combined sector *NAICS*.

7. CONCLUSION

This paper provided an overview of the sample design and estimation method that is planned and currently being tested for JOLTS. It raised issues regarding sample coordination and the government wide conversion to industry classification based on NAICS. We will evaluate methods and procedures as we collect JOLTS data and produce estimates for internal BLS review. BLS intends to publish experimental series on job openings and labor turnover in 2001.

References

Abraham, Katharine G. "Structural/Frictional vs. Deficient Demand Unemployment: Some New Evidence," *American Economic Review*, September 1983 73(4), pp. 708-724.

Butani, Shail, George Stamas, and Michael Brick, "Sample Redesign for the Current Employment Statistics Survey," ASA Papers and Proceedings, August 1997.

Chen, Jiahua, and Jun Shao, "Jackknife Variance Estimation For Nearest Neighbor Imputation," ASA Papers and Proceedings, August 1999.

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

Cochran, William G., *Sampling Techniques*, (1977), New York, John Wiley & Sons.

Goldenberg, Karen and Mary Anne Phillips, "Now that the Study is Over, What Did It Really Tell Us?..." ICES II Papers and Proceedings, June 2000.

Levin, K., C. Gimbel, T. Hagerty, S. Heltemes, A. Becker, and D. Cantor. *Job Openings and Labor Turnover Statistics (JOLTS) Pilot Study: Final Report*, Rockville, MD: Westat, January 2000.

Mikkelson, Gordon, Teresa Morisi and George Stamas, "Implementing the NAICS for Business Surveys at the BLS," ICES II, June 2000.

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 Statistic," ASA Papers and Proceedings, August 2000.

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