SUPPORTING STATEMENT FOR JOB OPENINGS AND LABOR TURNOVER SURVEY

OMB CONTROL NO. 1220-0170

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

For detailed technical materials on the sample allocation, selection, and estimation methods as well as other related statistical procedures see the BLS Handbook, internal BLS technical reports, and ASA papers listed in the references section of this document. The following is a brief summary of the primary statistical features of JOLTS.

1a. Universe

The Job Openings and Labor Turnover Survey measures the job openings, hires, total separations, quits, layoffs and discharges, and other separations for each month at the national level from a sample of about 20,700 establishments (worksites). The universe for this survey is the Quarterly Contribution Reports (QCR) filed by employers subject to State Unemployment Insurance (UI) laws. The U.S. Bureau of Labor Statistics (BLS) receives these QCR for the Quarterly Census of Employment and Wages (QCEW) program from the 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. The QCEW data, which are compiled for each calendar quarter, provide a comprehensive business name and address file with employment, wage, detailed geography (i.e., county), and industry information at the six-digit North American Industry Classification System (NAICS) level. This information is provided for over 8.5 million business establishments of which about 8.1 million are in the scope of this survey. Similar data for Federal Government employees covered by the Unemployment Compensation for Federal Employees program (UCFE) are also included. The final data is stored in a Longitudinal Data Base (LDB), which is then used as a sample frame for sample selection. Another data source used for sampling is the universe of railroad establishments obtained from the Federal Railroad Administration.

1b. Sample

Scope—The JOLTS sample is selected from the populations stated above excluding Puerto Rico and the Virgin Islands. It also excludes from the universe records that are for private household workers (NAICS 814110) and records from Agriculture, Forestry, Fishing and Hunting (NAICS 11) other than logging (113310). Records with average employment of zero in the last twelve months are also excluded from the universe.

Stratification—The JOLTS sample has about 20,700 establishments allocated based on the stratification of four census regions, 20 two-digit industry codes, and six employment size classes, including certainty establishments which have a certain level of employment, or the number of establishments in the universe for a sampling cell is less than or equal to 24. These certainty establishments are assigned a sampling weight of 1.00 and other establishments are assigned the sampling weight of the strata population count divided by the strata sample count. The population and sample counts and their employment levels by industry are shown in Table 1.

In addition to the annual sample, BLS added about 250 establishments in each of the three remaining quarters to represent newly formed businesses. The total sample size, therefore, is about 21,200 establishments. However, with a new sample selection every 1st quarter, the sample size is reduced to about 20,700 after discarding the out of business units not on the current frame.

Table 1: Distribution of Sample by Industry, April 2021

Industry	IDNAICS	Popn(N)	Popn(Emp)	Sample(n)	Sample(Empl)
Natural resources and mining	21	34,425	542,074	681	170,630
Construction	23	757258	7,077,086	968	265,881
Nondurable goods	31	126,062	4,466,433	852	434,741
Durable goods	33	209,132	7,368,935	960	980,262
Wholesale Trade	42	549,739	5,452,055	813	304,167
Retail trade	44	975,698	14,557,044	1,347	412,311
Transportation, warehousing, and utilities	48	265,449	6,065,338	1,117	1,261,909
Information	51	183,955	2,617,276	989	785,429
Finance and Insurance	52	488,403	5,982,796	838	518,816
Real estate and rental and leasing	53	402,310	2,090,698	870	189,211
Professional and business services	54	1,682,823	16,775,373	1,423	591,541
Employment services	56	100,170	3,089,671	940	600,979
Educational services	61	120,441	2,646,439	977	1,119,622
Health care and social assistance	62	975,158	18,587,304	1,406	1,742,137
Arts, entertainment, and recreation	71	136,223	1,532,516	898	639,019
Accommodation and food services	72	669,531	1,532,516	1,548	476,711
Other services	81	569,666	3,558,528	774	148,011
Federal government	91	28,090	2,259,581	882	1,104,289
State and local government education	92	72,994	9,163,878	1,124	3,104,216
State and local government non-education	93	160,969	8,776,692	946	1,664,433
Total annual sample including certainty/birth units		8,508,496	132,845,661	20,353	16,514,315
Quarterly sample of newly formed businesses				227	

2. Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**

- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

2a. Sample Design

Allocation method—The JOLTS sample design is a probability-based stratified random sample. The basic sampling unit is an establishment or worksite which generally remains in the survey for 36 months for a noncertainty establishment and stays out of the survey for the next three years after completion of the 36 months. Important features of the sample design are the use of stratified random sampling, a Neyman allocation (Cochran, 1977, pp. 259-261), and ratio estimators. The characteristics used to stratify the sample are geographic area by four census regions, 2-digit industry divisions as defined in Table 1, and six establishment employment size classes.

JOLTS characteristics are highly correlated with an establishment's employment level. Thus for a fixed sample size, stratified sampling results in a greater precision than simple random sampling. Given a fixed sample size, the Neyman allocation provides the maximum precision of an estimate. Some establishments are included in the sample with certainty.

Sample Rotation—The sample is divided into one certainty panel (panel 0) and 24 noncertainty panels. Each month, the oldest panel is rotated out and replaced by a new panel. Each panel is asked to provide data for 36 months. This maintains 24 active noncertainty panels for estimation.

In April 2009, new sampling procedures were implemented. During the annual sample, the previously sampled establishments still used in JOLTS estimation were updated, removing out-of-business establishments and updating industry and employment size class information. Also an age variable was added to all establishments in the sample. All the establishments to be used in the JOLTS estimation during the course of the sampling year were then weighted to the current sampling frame, so that they may represent the most current data. During that same sampling year, a quarterly birth sample was also implemented. The purpose of this birth sample is to enroll younger establishments into the JOLTS sample as soon as possible.

2b. Estimation Procedure

The survey utilizes a ratio estimator to improve the precision of the sample estimates. This estimator improves the precision of the sample estimates by utilizing the correlation between the employment data and the characteristics to be measured. A Horvitz-Thompson estimator (Lohr, 1999, Chapter 6.) with a ratio adjustment is used to produce estimates of surveyed characteristics at several levels of geographic and industrial detail. These estimates include the following:

- Totals
- Rates
- Estimates of monthly change

The generalized formula for totals for all survey characteristics (job openings, hires, etc.) for time period t is as follows for ready reference:

$$\hat{X}_t = \sum_{i \in cell} (W_{t,i} * NRAF_{t,cell} * BMF_{t,cell}) * X_{t,i}$$

where,

$X_{t,i}$ is the characteristic of interest for the i^th^ unit at time t.

$\hat{X}_t$ is the estimate of a characteristic at time t.

$W_{t,i}$ is the sample weight for the i^th^ unit at time t.

$NRAF_{t,cell}$ is the cell (Region/2-digit NAICS/SZC) non-response adjustment factor defined by ($\sum_{t,cell} \frac{W_{t,eligibles}}{W_{t,respondents}}$) at time t.

Where “respondents” are the all units reporting employment at time t and “eligible” are all sampled units excluding out-of-business units at time t within a cell.

BMF is the (Current Employment Statistics) Benchmark factor at time t. It is computed for each estimation cell as:

Benchmark factor = ($\frac{CES_{Emp_t}}{JOLTS_{Emp_t}}$)

where, CES_Emp_t is the employment level at time t obtained from the monthly Current Employment Statistics (CES) Survey, also known as the monthly Payroll Survey. The CES employment serves as a population control for each estimation cell and $JOLTS\_Emp_t$ is the sample weighted employment at time t.

The formula for the Job Openings rate is as follows:

$$JO\_Rate_t = \frac{\hat{JO}_t}{CES\_Emp_t + \hat{JO}_t}$$

where, $\hat{JO}_t$ is the estimated level of job openings at time t.

The generalized formula for all other rates is as follows:

$Rate_t = \frac{\hat{X}_t}{CES\_Emp_t}$ where $\hat{X}_t$ is the estimate of the characteristic at time t.

Details of JOLTS estimation are available at
https://www.bls.gov/opub/hom/jlt/calculation.htm#estimating-procedures

Birth/Death Model—As with any sample survey, the JOLTS sample can be only as current as its sampling frame. The time lag from the birth of an establishment until its appearance on the sampling frame is approximately one year. In addition, many of these new units may fail within the first year. Since these universe units cannot be reflected on the sampling frame immediately, the JOLTS sample cannot capture job openings, hires, and separations from these units during their early existence. BLS has developed a model to estimate birth/death activity using QCEW longitudinal data. The birth/death model also uses historical JOLTS data to estimate the total amount of "churn" (hires and separations) that exists in establishments of various sizes. The model then combines the estimated total churn with the QCEW employment change of younger units (less than 18 months) to estimate the number of hires and separations taking place in these units that cannot be measured through sampling.

The model-based estimate of total separations is distributed to the three components – quits; layoffs and discharges; and other separations – in proportion to their contribution to the sample-based estimate of total separations. Additionally, job openings for the modeled units are estimated by computing the ratio of openings to hires in the collected data and applying that ratio to the modeled hires. The estimates of job openings, hires, and separations produced by the birth/death model are then added to the sample-based estimates produced from the survey to arrive at the estimates for openings, hires, and separations. The derivation of the parameter's computational procedure is given in Appendix A and Appendix B.

Alignment—JOLTS hires minus separations should be comparable to the CES net employment change. The CES series is considered a highly accurate measure of net employment change owing to its very large sample size and annual benchmarking to universe counts of employment from the QCEW program. However, definitional differences as well as sampling and non-sampling errors between the two surveys historically caused JOLTS to diverge from CES over time. To limit the divergence and to improve the quality of the JOLTS hires and separations series, BLS implemented a monthly alignment method. This monthly alignment method applies the seasonally adjusted CES employment trends to the seasonally adjusted JOLTS implied employment trend (hires minus separations) forcing them to be approximately the same, while preserving the seasonality of the JOLTS data. A brief description is as follows.

First, the two series are seasonally adjusted and the difference between the JOLTS implied employment trend and the CES net employment change is calculated. Next, the JOLTS implied employment trend is adjusted to equal the CES net employment change through a proportional adjustment. This proportional adjustment procedure adjusts the two components (hires, separations) proportionally to their contribution to the total churn (hires plus separations). For example, if hires are 40 percent of the churn for a given month, they will receive 40 percent of the needed adjustment and separations will receive 60 percent of the needed adjustment. The following example illustrates the adjustment.

Example: let hires = 40; seps = 60; change in ces emp = -25
1) D = (hires - seps) - change of cesemp = 40 - 60 - (-25) = 5

2) PropAdj_Hires = hires / (hires + seps) * D = 40 / (40 + 60) * 5 = 2
3) PropAdj_Seps = seps / (hires + seps) * D = 60 / (40 + 60) * 5 = 3
4) Hires_sa = Hires - PropAdj_Hires = 40 - 2 = 38
5) Seps_sa = Seps - PropAdj_Seps = 60+3 = 63

Job openings are adjusted based on the adjustment made to hires. This adjustment applies the ratio of job openings to hires to the hires adjustment to arrive at the job openings adjustment. The adjusted job openings, hires, and separations are converted back to not seasonally adjusted data by reversing the application of the original seasonal factors. After the monthly alignment method is used to adjust the not seasonally adjusted level estimates, rate estimates are computed from the adjusted levels. The monthly alignment procedure assures a close match of the JOLTS implied employment trend with the CES trend for not seasonally adjusted data. The adjusted estimates are then again seasonally adjusted (see https://www.bls.gov/opub/hom/jlt/calculation.htm#seasonal-adjustment).

2c. Reliability

This survey is designed to produce reliable estimates of the characteristics of interest. For the period January 2021 through December 2021, the average relative standard errors for national estimates of job openings; hires; quits; layoffs and discharges; other separations; and total separations rate, respectively, were 1.97, 2.06, 2.19, 6.08, 6.65, and 2.32 percent. (See table 2.) Table 2 details the Average Relative Standard Error (RSE) of the JOLTS rates in percentages (Coefficients of Variation) for each JOLTS industry and variable. For the most part, the average JOLTS relative standard errors are below a reasonable acceptable RSE (that is, an RSE of 30 percent). The notable exception would be other separations. The mean Other Separations rate is generally near zero and this proximity of the mean to zero can substantially, on occasion, raise the RSE above a reasonable RSE.

The estimation of sample variances for the JOLTS survey is accomplished through the method of Balanced Half Samples (BHS) similar to CES. This replication technique uses half samples of the original sample and calculates estimates using those subsamples. The replicate weights in both half samples are modified using Fay's method of perturbation. The sample variance is calculated by measuring the variability of the estimates made from these subsamples. (For a detailed mathematical presentation of this method, see Handbook of Methods, BLS Chapter 2, Bureau of Labor Statistics, 2011 or http://www.bls.gov/opub/hom/homch2.htm under Reliability of Estimates.)

We compute the replicate estimates $\hat{Y}^{(\alpha)}$ using the whole sample rather than only half of the sample, as with the original BRR method. For each replicate, sample units are used with the modified weights $w_i^{(\alpha)}$:

$$w_i^{(\alpha)} = \left(1 + \gamma G_i h^{(\alpha)} \sqrt{1-f}\right) w_i$$

Where

$\gamma = 0.5$, a perturbation factor

$G_i = \pm 1$, the replicate indicator

$h(\alpha)$ is the element of the Hadamard matrix (row α for a given column)

f is the sampling fraction

w_i is the selection weight

In the above formula, the factor $\sqrt{1-f}$ is not part of the Fay's procedure, it was added to account for sampling from the finite population.

After we obtain the replicate estimates, we compute the variance using the usual formula:

$$Var_{FayBRR}\left(\hat{Y}\right) = \frac{1}{A\gamma^2}\sum_{\alpha=1}^{A}\left(\hat{Y}^{(\alpha)} - \hat{Y}\right)^2$$

where A is the number of replicates, 114 for JOLTS from a Hadamard matrix of order 116.

Before estimates of these characteristics are released to the public, they are first screened to ensure that they do not violate the Bureau of Labor Statistics' (BLS) confidentiality pledge. A promise is made by the BLS to each respondent that BLS will not release its reported data to the public in a manner which would allow others to identify the establishment, firm, or enterprise. Estimates which fail confidentiality screening based on p-percent rule for disclosure (see Federal Committee on Statistical Methodology Working paper 22) are not published.

2d. Revisions

In order to reflect revisions in the CES (Current Employment Statistics) estimates, the CES second closing revision is reflected in the JOLTS second closing estimates, and the final CES revision is incorporated in the JOLTS estimates on a yearly basis when JOLTS rebenchmarks to CES after CES estimates are benchmarked against the QCEW population.

2e. Specialized Procedures

BLS conducted extensive research to: (1) improve sampling procedures to bring in birth units on a timely basis in order to reduce bias; (2) improve the quality of the reported data in order to reduce response error; and (3) improve data collection procedures in order to increase response rates. The current respondent annual burden is about 17,326 hours with the current JOLTS unweighted response rate of 41.9. This calculation is derived as:

Annual burden hours = 20,700 X 0.419 X 12 X 10 / 60 = approximately 17,326 hours.

Where, 20,700 is the total number of establishments in the current annual sample and 100 establishments for each of the three remaining quarters for births; 0.419 is the target goal of response rate; 12 months; 10 minutes per schedule; and 60 minutes. NOTE: The actual burden will vary depending on how many birth establishments are sampled during the whole year.

2f. Data Collection Cycles

JOLTS data are collected every month.

Table 2	Average Relative Standard Error in Percentages (Coefficients of Variation) for Rates by Industry January 2021 - December 2021						
ID	**Industry/Rates**	**Job Openings**	**Hires**	**Quits**	**Layoffs and Discharges**	**Other Separations**	**Total Separations**
TOT	Total	1.97	2.06	2.19	6.08	6.65	2.32
PRI	Total Private	2.14	2.16	2.29	6.31	7.93	2.44
21	Mining and Logging	11.60	13.32	13.79	24.26	46.85	12.33
23	Construction	11.70	7.93	9.52	18.11	38.45	9.37
MFG	Manufacturing	5.67	6.27	6.42	13.83	13.29	6.45
DUR	Durable Goods	7.07	6.82	7.90	17.79	17.06	7.61
NDR	Nondurable Goods	8.61	10.18	10.08	18.61	18.74	10.53
TTU	Trade, Transportation, and Utilities	5.30	4.49	5.02	12.59	18.64	4.45
42	Wholesale Trade	9.34	8.82	11.00	31.03	41.70	14.36
44	Retail Trade	8.41	5.97	6.37	14.17	27.26	5.27
48	Transportation, Warehousing, and Utilities	6.66	8.19	9.28	23.36	23.77	8.58
51	Information	15.54	14.61	14.56	37.08	30.94	15.99
FIR	Financial Activities	8.14	7.67	10.94	21.27	26.36	10.64
52	Finance and Insurance	9.69	9.87	13.71	23.12	27.52	12.96
53	Real Estate and Rental and Leasing	13.91	10.59	14.95	31.61	55.59	16.22
54	Professional and Business Services	4.81	5.34	5.17	12.49	12.95	6.12
55	Employment Services	7.05	9.26	8.47	18.56	21.45	8.63
EHS	Education and Health Services	4.32	3.65	4.88	12.48	16.04	4.48
61	Educational Services	8.42	7.31	8.81	22.83	32.32	9.39
62	Health Care and Social Assistance	4.65	4.00	5.28	13.78	17.39	4.85
L&H	Leisure and Hospitality	5.17	4.17	4.37	15.98	21.81	4.85
71	Arts, Entertainment, and Recreation	8.22	9.60	8.86	24.31	31.99	11.74
72	Accommodation and Food Services	5.70	4.47	4.61	18.74	24.42	5.18
81	Other Services	10.00	13.30	17.57	35.99	56.01	16.77
GOV	Government	3.48	4.13	4.79	12.16	6.77	4.47
91	Federal	7.08	5.56	7.36	10.42	9.60	6.17
S&L	State and Local	3.89	4.60	5.31	13.61	8.25	5.05
SLE	State and Local Education	4.59	5.61	6.55	16.83	11.43	6.52
SLN	State and Local, excluding Education	5.40	5.96	7.88	15.54	10.58	6.75

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

3a. Maximize Response Rates

To maximize the response rate for this survey, interviewers initially refine addresses ensuring appropriate contact with the employer. Then, employers are mailed a folder containing a JOLTS brochure and data collection form, along with a cover letter explaining the importance of the survey and the need for voluntary cooperation, and pledging confidentiality. An interviewer calls the establishment after the package is sent and attempts to enroll them into the survey. Nonrespondents and establishments that are reluctant to participate are recontacted by an interviewer specially trained in refusal aversion and conversion. The response rates for April 2022 are shown below in Table 3.

3b. Nonresponse Adjustment

As with other surveys, JOLTS experiences a certain level of nonresponse. To adjust for the nonresponses, JOLTS has divided the nonresponse into two groups: (1) unit nonrespondents and failure to enroll; and (2) item nonresponse. Unit nonrespondents are the establishments that do not report the employment and item non-respondents are the establishments that do report employment but do not report one or more data items, for example, job openings or hires.

The unit nonresponse is treated using a Nonresponse Adjustment Factor (NRAF) as explained in the estimation procedure section of this document and item non-response is adjusted using item imputation. Within each sampling cell, NRAFs are calculated every month based on the ratio of the number of viable establishments to the number of usable respondents in that month. The details regarding the NRAF procedure are given in https://www.bls.gov/opub/hom/jlt/calculation.htm#unit-nonresponse-adjustment. The method used for item imputation is item nonresponse adjustment. Details of this procedure are available at https://www.bls.gov/opub/hom/jlt/calculation.htm#item-nonresponse-adjustment. New methods are being developed to improve on the current item nonresponse imputation.

3c. Nonresponse Bias Assessment and Research

As mentioned earlier, JOLTS has developed a birth/death model of hires and separations based on historical QCEW-LDB data. The model allows for establishment-level estimates of hires and separations for all establishments on the QCEW-LDB. Since the QCEW-LDB serves as the sampling frame for JOLTS, it is possible to produce model hires and separations estimates for all establishments sampled by JOLTS. Consequently, it is possible to compare the model estimates for respondents to non-respondents for establishments in the JOLTS sample. The research indicates the JOLTS respondents differ from non-respondents in one important aspect: the rate of out-of-business establishments for responding sample is much lower than for nonresponding sample. That is, it appeared that establishments exiting the labor force were not likely to report JOLTS data as they exited. The JOLTS birth/death model has been added to the estimation process in an attempt to mitigate this bias.

Table 3: Unweighted and Weighted Unit Response Rates by Industry, April 2022

Industry Division	Sampled (n)	Respondents (n)	Out of Business Respondents	Unweighted Response Rate %	Weighted Response Rate %
1.0 Total	20,353	7,338	580	37%	45%
2.0 Total Private	17,401	5,946	514	35%	34%
2.1 Mining and Logging	681	295	33	39%	49%
2.2 Construction	968	373	26	40%	44%
2.3 Manufacturing	1,812	654	44	37%	49%
2.3.1 Durable Goods	960	356	22	38%	50%
2.3.2 Nondurable Goods	852	298	22	36%	46%
2.4 Transportation, Warehousing, and Utilities	3,277	1,108	92	35%	47%
2.4.1 Wholesale Trade	813	295	17	37%	51%
2.4.2 Retail Trade	1,347	440	32	33%	45%
2.4.3 Transportation, Warehousing, and Utilities	1,117	373	43	35%	46%
2.5 Information	989	228	29	24%	34%
2.6 Financial Activities	1,708	532	44	32%	44%
2.6.1 Finance and Insurance	838	253	21	31%	46%
2.6.2 Real Estate and Rental and Leasing	870	279	23	33%	43%
2.7 Professional and Business Services	1,423	492	41	36%	48%
2.8 Employment Services	940	246	22	38%	42%
2.9 Educational and Health Services	2,383	919	75	40%	43%
2.9.1 Educational Services	977	390	20	41%	43%
2.9.2 Health Care and Social Assistance	1,406	529	55	39%	43%
2.10 Leisure and Hospitality	2,446	849	91	36%	41%
2.10.1 Arts, Entertainment, and Recreation	898	342	25	39%	47%
2.10.2 Accommodation and Food Services	1,548	507	66	34%	39%
2.11 Other Services	774	292	17	39%	47%
3.0 Government	2,952	1392	66	48%	52%
3.1 Federal	882	387	21	45%	42%
3.2 State and Local	2,070	1005	45	50%	54%
3.2.1 State and Local Education	1,124	520	31	48%	46%
3.2.2 State and Local, excluding Education	946	485	14	52%	57%

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

The initial survey's questionnaire was developed and tested using cognitive design techniques. The questionnaire has been used in production of estimates from December 2000 to the present. A Response Analysis Survey (RAS) was conducted on two major industries—Temporary Help Services and State and Local Government Education—to assess the sources of divergence between the employment change from CES and the implied employment change from JOLTS hires minus separations. In the former industry, businesses have a difficult time reporting hires and separations of temporary help workers. In the latter industry, employers have difficulty reporting hires and separations of student workers. BLS now devotes additional resources to the collection, editing, and review of data for these industries. BLS analysts more closely examine reported data that do not provide a consistent picture over time, and recontact the respondents as necessary. Analysts work with the respondents to adjust their reporting practices as possible. Units that cannot be reconciled but are clearly incorrect on a consistent basis are not used; they are replaced by imputed values using standard techniques.

Periodic tests similar to the RAS are necessary to understand the quality of the reported data and to improve the process in order to reduce sources of error or bias. In the future, BLS may make a nonsubstantive change request of approximately 400 respondent burden hours for future cognitive tests, such as a response analysis survey on the reporting of data items. The questionnaire(s) as well as relevant materials will be provided to OMB at the time of the request.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze person(s) who will actually collect and/or analyze the information for the agency.

Mr. Edwin Robison, Acting Chief, Statistical Methods Division of the Office of Employment and Unemployment Statistics, is responsible for the statistical aspects of the JOLTS program. Mr. Robison can be reached on 202-691-6363. As mentioned in the above paragraph, BLS seeks consultation with other outside experts on an as needed basis.

References

Bureau of Labor Statistics "Handbook of Methods," Chapter 2, Bureau of Labor Statistics, 2004, http://www.bls.gov/opub/hom/homch2.htm.

Cochran, William, G., (1977), Sampling Techniques 3rd Ed., New York, Wiley and Sons, 98, 259-261.

Crankshaw, Mark (April-2008), BLS Comparing the Level of Employment Churn: JOLTS Respondents vs. JOLTS Non-Respondents (attached).

Crankshaw, Mark (July, 2008), Simulating JOLTS Hires and Separations Data Using Historical QCEW Data, http://www.bls.gov/osmr/pdf/st100140.pdf. (paper attached)

Crankshaw, Mark and Stamas, George (Aug, 1999), "Sample Design in the Job Openings and Labor Turnover Survey," ASA Papers and Proceedings, http://www.bls.gov/osmr/pdf/st000140.pdf.

Federal Committee on Statistical Methodology, Subcommittee on Disclosure Limitation Methodology, "Statistical Policy Working Paper 22," http://www.fcsm.gov/working-papers/SPWP22_rev.pdf.

Goodale, Sarah (July-2008), BLS Internal Document: Adjusting Sampling Weights in The JOLTS Sample (working paper-attached).

Goodale, Sarah and Greene, Darrell (Oct 2009), "Improving the Job Openings and Labor Turnover Survey's Sampling Procedure," ASA Papers and Proceedings, http://www.bls.gov/osmr/pdf/st090040.pdf.

Job Openings and Labor Turnover Statistics survey: Statistical Methods Documentation (2006).

Lohr, Sharon L. (1999), Sampling: Design and Analysis, Brooks/Cole Publishing Company, Chapter 6.

Mueller, Kirk; Stamas, George; and Butani, Shail, "Nonresponse Adjustment In Certainty Strata For An Establishment Survey," http://www.bls.gov/osmr/pdf/st950130.pdf.

Cheng, Edmond; Hudson, Nicole; Kropf, Jurgen; and Mercurio, Jeanine, "The CES/JOLTS Divergence: How to Apply the Monthly Alignment Method to help close the Gap," http://www.bls.gov/osmr/pdf/st090300.pdf.

Appendix A

Deriving Ω levels

The initial approach taken to estimate Ω was to utilize the hires and separations rates for stable, contracting, and expanding units. For each type of unit (stable, contracting, and expanding) the hires and separations data was available at the industry level and size level but not for a combined industry-size level. An approximation was made using industry level data and increasing or decreasing Ω levels for each size class within the industry.

Another approach has been taken. In this approach, a dataset containing JOLTS respondent data from Dec 2000 to April 2008 was created such that all reporters reported two consecutive months of data (a necessary precondition for simulation) and, additionally, all reporters reported both hires and separations. Using this data set, a crude simulation was made such that:

1. For stable units, the hires and separations rates found on page 3 were utilized. The rates were smoothed so that the hires rate equaled the separations rate and the industry-size estimate was made using the initial approach.
2. For expanding units, the hires were set equal to the increase in employment and the separations were set to zero.
3. For contracting units, the separations were set equal to the absolute decrease in employment and the hires were set to zero.

This crude simulation would measure the amount of net churn for a given industry-size cell. Comparing this estimate with the actual reported values would enable one to solve for the underlying churn (and hence Ω level) for all industry-size cells. The difference between the reported value and the net churn is equal to the underlying churn (that is, the hires and separations reported in addition to the net change in employment).

Following is an example to illustrate the technique used to derive Ω levels:

ID: 21 (Mining & Natural resources)
Size: 4 (250-999 employees)

<u>Reported Data</u>

Employment:	1,258,767
Hires:	30,277
Separations:	28,652

Crude Simulated Hires:	19,799	Reported – Simulated:	10,478
Crude Simulated Separations:	16,802	Reported – Simulated:	11,850

$$\Omega_h = 10{,}478/1{,}258{,}767 = 0.83\ \%$$

Ω_s = 11,850/1,258,767 = 0.94 %

Below are the calculated Ω levels for each industry size:

ID	S	Emp	Orig_Hires	Orig_Seps	C_Impied	C_Implied	Ω_h	Ω_s
21	1	14504	594	648	511	601	0.57%	0.32%
21	2	80094	3081	3094	2424	2252	0.82%	1.05%
21	3	314440	10471	10491	5884	5814	1.46%	1.49%
21	4	1258767	30277	28652	19799	16802	0.83%	0.94%
21	5	1492912	23759	25011	15333	15718	0.56%	0.62%
21	6	29894	366	140	513	449	0.00%	0.00%
23	1	22379	1004	1047	859	809	0.65%	1.06%
23	2	229794	11431	12046	8913	8947	1.10%	1.35%
23	3	569558	32932	32755	20682	21724	2.15%	1.94%
23	4	637288	40631	41288	20984	23587	3.08%	2.78%
23	5	1141391	84884	66046	28781	38916	4.92%	2.38%
23	6	225161	5200	6448	2863	4657	1.04%	0.80%
31	1	27293	493	706	402	604	0.33%	0.37%
31	2	98963	3691	3854	2629	2730	1.07%	1.14%
31	3	971022	25611	28187	16726	19297	0.92%	0.92%
31	4	3160271	67644	82306	31846	45164	1.13%	1.18%
31	5	4188433	99473	108788	40411	53721	1.41%	1.31%
31	6	942869	13428	13310	4075	4996	0.99%	0.88%
33	1	15587	513	576	457	465	0.36%	0.71%
33	2	176236	5770	6437	3925	4154	1.05%	1.30%
33	3	1743483	45242	50697	26280	29533	1.09%	1.21%
33	4	5784226	109121	137136	55017	75608	0.94%	1.06%
33	5	7865330	104141	124366	62973	80668	0.52%	0.56%
33	6	16593811	117190	154901	62815	102656	0.33%	0.31%
42	1	28627	595	704	705	582	0.00%	0.43%
42	2	243266	5724	5763	4829	4138	0.37%	0.67%
42	3	669718	17239	17275	10116	17998	1.06%	0.00%
42	4	928400	22793	25240	9943	13239	1.38%	1.29%
42	5	3191628	49358	58455	21868	29618	0.86%	0.90%
44	1	103130	2655	3048	2828	2016	0.00%	1.00%

44	2	365482	15386	15776	10321	11012	1.39%	1.30%
44	3	1787337	91017	87324	41122	44166	2.79%	2.41%
44	4	1462379	78863	76591	32418	34178	3.18%	2.90%
44	5	2319000	121224	112461	58725	33591	2.70%	3.40%
44	6	373621	39839	48516	20860	20363	5.08%	7.54%
48	1	13379	375	458	325	341	0.37%	0.87%
48	2	320310	4361	4977	5531	5423	0.00%	0.00%
48	3	1182664	21597	20777	17483	11835	0.35%	0.76%
48	4	1218587	39959	40399	18051	19338	1.80%	1.73%
48	5	6260773	130071	110621	51525	52898	1.25%	0.92%
48	6	13594810	202640	209447	82383	99945	0.88%	0.81%
51	1	8418	200	212	175	167	0.30%	0.53%
51	2	87343	2472	2624	1554	1522	1.05%	1.26%
51	3	308407	8081	8687	5336	6222	0.89%	0.80%
51	4	838610	14833	18489	9165	11061	0.68%	0.89%
51	5	1886036	30291	33930	12912	19433	0.92%	0.77%
51	6	1855299	44643	47192	11134	10573	1.81%	1.97%
52	1	17147	321	334	358	337	0.00%	0.00%
52	2	166324	3600	3698	2855	3059	0.45%	0.38%
52	3	641224	14313	14958	8151	8336	0.96%	1.03%
52	4	2358154	49490	45860	26845	19376	0.96%	1.12%
52	5	3832948	64972	73486	28153	37994	0.96%	0.93%
52	6	6305608	97040	106120	28060	21219	1.09%	1.35%

ID	**S**	**Emp**	**Orig_Hires**	**Orig_Seps**	**C_Impied**	**C_Implied**	Ω_h	Ω_s
53	1	13050	428	432	315	336	0.87%	0.74%
53	2	62884	2041	2101	1411	1609	1.00%	0.78%
53	3	164801	7417	7419	4436	4763	1.81%	1.61%
53	4	609318	24652	23280	10965	9799	2.25%	2.21%
53	5	249919	13884	16235	8279	10141	2.24%	2.44%
54	1	46235	1360	1556	1108	1341	0.55%	0.47%
54	2	313919	10425	10683	8440	8186	0.63%	0.80%
54	3	1821071	75467	68074	41635	47733	1.86%	1.12%
54	4	3695469	153032	137626	67723	67909	2.31%	1.89%
54	5	5082319	122746	133124	52634	72164	1.38%	1.20%
5	6	10917197	174524	194316	97378	94517	0.71%	0.91%

4								
56	1	3562	135	146	148	138	0.00%	0.22%
56	2	11907	883	760	547	1054	2.82%	0.00%
56	3	49219	8355	6761	2703	2491	11.48%	8.68%
56	4	71476	9800	8305	3433	4859	8.91%	4.82%
56	5	211376	19185	17350	5540	6550	6.46%	5.11%
56	6	1105696	116331	108415	12693	14129	9.37%	8.53%
61	1	19363	366	323	380	351	0.00%	0.00%
61	2	73520	2055	1745	2044	1812	0.01%	0.00%
61	3	298031	8374	6989	7330	6292	0.35%	0.23%
61	4	1048565	25114	21040	27659	26735	0.00%	0.00%
61	5	3790949	71597	64193	77364	67090	0.00%	0.00%
61	6	9204829	143095	103710	127526	123120	0.17%	0.00%
62	1	43209	1410	1356	1008	934	0.93%	0.98%
62	2	308251	9434	9047	6529	6021	0.94%	0.98%
62	3	2094016	72128	63981	31118	33679	1.96%	1.45%
62	4	4689700	125028	101436	49086	41601	1.62%	1.28%
62	5	23037096	416536	314362	149685	126420	1.16%	0.82%
62	6	29557101	458090	326803	153575	74406	1.03%	0.85%
71	1	8879	399	413	610	587	0.00%	0.00%
71	2	52249	3553	3290	2642	2682	1.74%	1.16%
71	3	204795	16241	14560	14019	13822	1.08%	0.36%
71	4	838029	61092	55543	46479	45513	1.74%	1.20%
71	5	3298756	270103	255808	166759	174070	3.13%	2.48%
71	6	299834	10546	8903	4326	4100	2.07%	1.60%
72	1	46600	1722	1525	1109	3739	1.32%	0.00%
72	2	393101	25409	25245	13450	13955	3.04%	2.87%
72	3	833812	54413	51148	25741	28916	3.44%	2.67%
72	4	919954	45839	43999	22205	25571	2.57%	2.00%
72	5	4091228	156671	152035	73334	77568	2.04%	1.82%
72	6	2349840	51009	45821	18157	18343	1.40%	1.17%
81	1	24837	591	708	501	562	0.36%	0.59%
81	2	124960	3852	4274	3084	3276	0.61%	0.80%
81	3	298374	14960	13703	8678	8376	2.11%	1.79%
81	4	684543	31511	27756	16681	18698	2.17%	1.32%

81	5	1256498	60786	41757	26287	24561	2.75%	1.37%
91	1	3194679	49421	45021	23179	11087	0.82%	1.06%
91	2	10318038	158249	128050	49293	37810	1.06%	0.87%
91	3	3055757	39851	38232	46782	17132	0.00%	0.69%
91	4	2656688	49661	43392	22031	12157	1.04%	1.18%
91	5	13969519	222103	200263	58426	63754	1.17%	0.98%
91	6	45700741	585870	454070	269931	166348	0.69%	0.63%
92	1	208970	2701	2438	5013	4114	0.00%	0.00%
92	2	379681	5221	4214	9235	7343	0.00%	0.00%
92	3	2252458	34030	26720	47121	42958	0.00%	0.00%
92	4	4586257	69058	48199	112835	102262	0.00%	0.00%
ID	**S**	**Emp**	**Orig_Hires**	**Orig_Seps**	**C_Impied**	**C_Implied**	Ω_h	**Resid_S%**
92	5	18518842	311435	207704	396109	356746	0.00%	0.00%
92	6	103749630	1880372	1508904	1480803	1404892	0.39%	0.10%
93	1	112428	2099	1485	1441	1637	0.59%	0.00%
93	2	494395	8066	7322	5725	5367	0.47%	0.40%
93	3	4712258	58616	50842	35485	35649	0.49%	0.32%
93	4	9238622	147199	132673	85143	90945	0.67%	0.45%
93	5	22185518	338298	287402	161892	144201	0.80%	0.65%
93	6	62082500	748715	591932	358277	298834	0.63%	0.47%

NOTE: Negative values were set to 0.00%
NOTE: For the simulation the Ω levels were rounded to the nearest tenth of a percentage point.

A simulation was performed on the JOLTS data and a comparison was made against the actual reported data. Here are the results:

ID	Type	N	Emp	Avg Emp	H	TS	HR	TSR	CR
21	Rep	14,153	3,190,611	225	68,548	68,036	2.15%	2.13%	4.28%
21	Sim	14,153	3,190,611	225	69,588	66,854	2.18%	2.10%	4.28%
23	Rep	33,114	2,825,571	85	176,082	159,630	6.23%	5.65%	11.88%
23	Sim	33,114	2,825,571	85	175,866	160,117	6.22%	5.67%	11.89%
31	Rep	30,963	9,388,851	303	210,340	237,151	2.24%	2.53%	4.77%
31	Sim	30,963	9,388,851	303	207,857	238,280	2.21%	2.54%	4.75%
33	Rep	52,305	32,178,673	615		474,083	1.19%	1.47%	2.66%

					381,977				
33	Sim	52,305	32,178,673	615	383,807	467,009	1.19%	1.45%	2.64%
42	Rep	28,141	5,061,639	180	95,709	107,437	1.89%	2.12%	4.01%
42	Sim	28,141	5,061,639	180	96,820	108,890	1.91%	2.15%	4.06%
44	Rep	62,609	6,410,949	102	348,984	343,716	5.44%	5.36%	10.80%
44	Sim	62,609	6,410,949	102	349,473	343,704	5.45%	5.36%	10.81%
48	Rep	21,943	22,590,523	1,030	399,003	386,679	1.77%	1.71%	3.48%
48	Sim	21,943	22,590,523	1,030	405,638	385,305	1.80%	1.71%	3.50%
51	Rep	12,190	4,984,113	409	100,520	111,134	2.02%	2.23%	4.25%
51	Sim	12,190	4,984,113	409	100,344	112,266	2.01%	2.25%	4.27%
52	Rep	22,861	13,321,405	583	229,736	244,456	1.72%	1.84%	3.56%
52	Sim	22,861	13,321,405	583	228,934	254,784	1.72%	1.91%	3.63%
53	Rep	12,557	1,099,972	88	48,422	49,467	4.40%	4.50%	8.90%
53	Sim	12,557	1,099,972	88	47,950	49,355	4.36%	4.49%	8.85%
54	Rep	57,411	21,876,210	381	537,554	545,379	2.46%	2.49%	4.95%
54	Sim	57,411	21,876,210	381	537,036	545,247	2.45%	2.49%	4.95%
ID	**Type**	**N**	**Emp**	**Avg Emp**	**H**	**TS**	**HR**	**TSR**	**CR**
56	Rep	2,764	1,453,236	526	154,689	141,737	10.64%	9.75%	20.40%
56	Sim	2,764	1,453,236	526	154,891	141,882	10.66%	9.76%	20.42%
61	Rep	15,046	14,435,257	959	250,601	198,000	1.74%	1.37%	3.11%
61	Sim	15,046	14,435,257	959	261,740	226,161	1.81%	1.57%	3.38%
62	Rep	64,890	59,729,373	920	1,082,626	816,985	1.81%	1.37%	3.18%
62	Sim	64,890	59,729,373	920	1,079,522	830,742	1.81%	1.39%	3.20%
71	Rep	14,377	4,702,542	327	361,934	338,517	7.70%	7.20%	14.90%
71	Sim	14,377	4,702,542	327	360,231	340,092	7.66%	7.23%	14.89%
72	Rep	43,329	8,634,535	199	335,063	319,773	3.88%	3.70%	7.58%
72	Sim	43,329	8,634,535	199	332,903	327,243	3.86%	3.79%	7.65%
81	Rep	23,447	2,389,212	102	111,700	88,198	4.68%	3.69%	8.37%
81	Sim	23,447	2,389,212	102	112,234	88,836	4.70%	3.72%	8.42%
91	Rep	10,739	78,895,422	7,347		909,028	1.40%	1.15%	2.55%

					1,105,155				
91	Sim	10,739	78,895,422	7,347	1,099,790	926,418	1.39%	1.17%	2.57%
92	Rep	46,938	129,695,838	2,763	2,302,817	1,798,179	1.78%	1.39%	3.16%
92	Sim	46,938	129,695,838	2,763	2,378,613	2,174,869	1.83%	1.68%	3.51%
93	Rep	53,067	99,925,721	1,883	1,302,993	1,071,656	1.30%	1.07%	2.38%
93	Sim	53,067	99,925,721	1,883	1,312,940	1,049,607	1.31%	1.05%	2.36%
ALL	Rep	622,844	522,789,653	839	9,604,453	8,409,241	1.84%	1.61%	3.45%
ALL	Sim	622,844	522,789,653	839	9,696,177	8,837,661	1.85%	1.69%	3.55%

Appendix B

The Current JOLTS imputation vs. the Simulation

A sample of the JOLTS dataset mentioned in Appendix A was drawn. The sample consisted of approximately 14% of the dataset. The units sampled received two treatments: (1) using the simulation, hires and separations data were produced; and (2) they had hires and separations data imputed using the current JOLTS imputation algorithm.

The current JOLTS imputation algorithm is a hot-deck nearest neighbor technique. The imputation cell (region/industry) is sorted by reported monthly employment. Units in need of imputation borrow from the closest available donor within the cell with respect to employment. Ties in closeness are broken randomly.

In this treatment we can directly compare the actual reported hires and separations directly against the hires and separations for the simulated and imputed data.

Below is a summary of the analysis:

ID	N	Emp	OHR	OSR	OCR	SHR	SSR	SCR	IHR	ISR	ICR
21	2,002	512,189	1.87%	1.84%	3.71%	1.84%	1.67%	3.51%	1.85%	1.74%	3.58%
23	4,353	374,430	5.70%	5.56%	11.26%	6.21%	5.40%	11.61%	5.29%	5.00%	10.29%
31	3,842	1,209,735	2.41%	2.44%	4.85%	2.16%	2.30%	4.46%	2.09%	2.16%	4.25%
33	6,898	4,207,648	1.25%	1.30%	2.55%	1.12%	1.22%	2.34%	1.26%	1.37%	2.64%
42	3,715	799,824	1.92%	1.96%	3.88%	2.11%	1.75%	3.86%	1.91%	1.80%	3.72%
44	8,293	848,729	5.13%	5.00%	10.13%	5.03%	4.79%	9.82%	4.55%	4.50%	9.04%
48	2,843	2,812,648	1.88%	1.84%	3.72%	1.77%	1.72%	3.48%	1.86%	1.87%	3.73%
51	1,595	709,821	2.13%	2.22%	4.35%	2.05%	2.07%	4.13%	1.87%	2.12%	3.99%
52	2,960	1,644,613	1.62%	1.65%	3.28%	1.74%	1.80%	3.54%	1.65%	1.57%	3.23%
53	1,700	152,889	4.26%	4.35%	8.61%	4.19%	4.62%	8.80%	4.15%	3.82%	7.98%
54	7,312	2,353,207	2.34%	2.06%	4.40%	2.31%	2.46%	4.77%	2.23%	2.18%	4.41%
56	421	346,873	10.67%	8.66%	19.33%	9.75%	9.83%	19.58%	6.02%	6.61%	12.63%
61	2,372	2,495,490	1.46%	1.33%	2.79%	1.74%	1.74%	3.47%	1.61%	1.40%	3.01%
62	9,064	9,680,562	1.71%	1.30%	3.00%	1.68%	1.37%	3.05%	1.67%	1.28%	2.96%
71	2,157	772,826	8.32%	6.55%	14.88%	7.81%	6.66%	14.47%	5.15%	5.40%	10.55%
72	5,763	1,328,931	3.82%	3.40%	7.22%	3.82%	3.49%	7.31%	3.47%	3.13%	6.60%
81	3,128	422,095	4.83%	3.52%	8.34%	4.65%	3.52%	8.16%	3.69%	3.12%	6.82%
91	1,445	10,690,793	1.38%	1.37%	2.75%	1.08%	1.22%	2.30%	1.47%	1.34%	2.82%
92	6,722	21,219,082	1.78%	1.47%	3.25%	1.78%	1.49%	3.27%	1.72%	1.31%	3.03%
93	7,532	16,074,949	1.23%	1.11%	2.34%	1.24%	1.06%	2.30%	1.28%	0.99%	2.27%
AL L	**84,117**	**78,657,334**	**1.81%**	**1.60%**	**3.41%**	**1.76%**	**1.60%**	**3.36%**	**1.74%**	**1.49%**	**3.23%**

Main Finding on the 1st randomly selected sample of reported JOLTS data:

The imputed values show less churn (both hires and especially separations) than do the actual and simulated values.

Reference- attachment.

Comparing the Level of Employment Churn:
JOLTS Respondents vs. JOLTS Non-Respondents
Mark Crankshaw
April 2008

Introduction

One assumption in the JOLTS survey is that the non-respondents to the JOLTS survey do not systematically differ from respondents. This assumption has been questioned by some and it has been asserted that the non-respondents to the JOLTS survey are more volatile with respect to monthly employment than are respondents to the survey; that is, employment churning of non-respondents greatly exceeds the employment churning of respondents. This would imply that the estimated rates of JOLTS variables that measure employment churning, namely hires and separations, are systematically biased in the downward direction.

One way to test the hypothesis that JOLTS non-respondents have greater employment churning than respondents is to match the JOLTS sample to the Longitudinal Database (LDB). The LDB contains historical employment data for all JOLTS records. The absolute month-to-month employment of matched units on the LDB can serve as a proxy for employment churning. Those establishments with a higher absolute average employment change on the LDB could be assumed to have greater levels of employment churn than establishments with lower absolute average employment change.

Making the Comparison

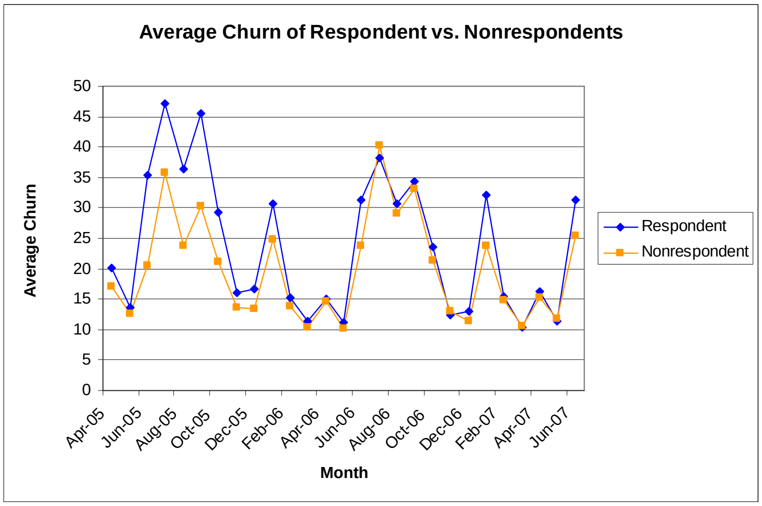
To test whether non-respondents have higher levels of average absolute employment change than respondents we can contrast the average absolute employment change on the LDB for all non-respondents to the JOLTS survey against the average absolute employment change on the LDB for all respondents to the JOLTS survey. If the average absolute employment change for non-respondents is statistically higher than the average absolute employment change for respondents, then the assumption of no difference is violated. However, if there is no statistical difference found, then the assertion that non-respondents systematically differ from the respondents is not backed up by LDB data.

All JOLTS records were matched with the LDB (over the period April, 2005 to June 2007). The absolute employment change was calculated for all matched records. The average absolute employment change for non-respondents and respondents was calculated.

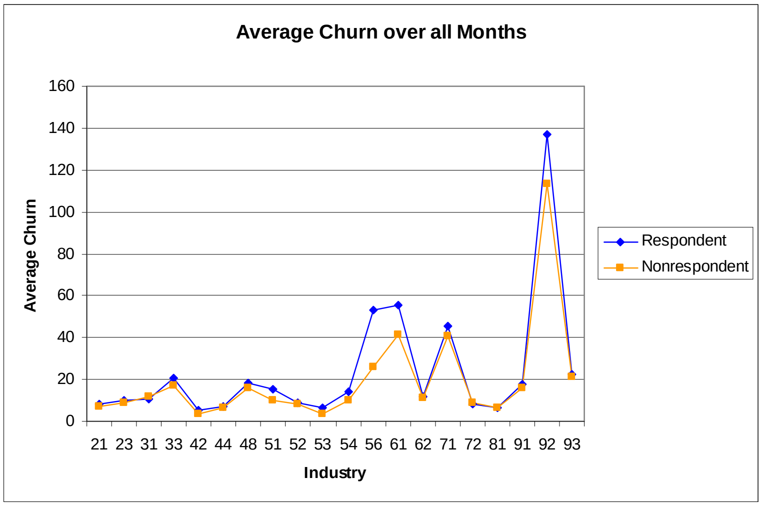
Findings

There was no evidence found to support the hypothesis that non-respondents systematically differ from respondents to the JOLTS survey. Overall, the average absolute employment change from month to month for respondents was 23.88, while for non-respondents it was 19.83. The difference between the two was not statistically significant. This finding of no difference was found across all months analyzed as well as across all industries.

The graph below charts the average absolute employment change across the months analyzed:



The graph below charts the average absolute employment change for all industries:



The table below details the comparison by industry:

Industry	**JOLTS ID**	**Absolute Average Emp change Respondents**	**Absolute Average Emp change Non-Respondents**
Natural Resources	21	8.35	7.18
Construction	23	10.06	8.89
Durable MFG	31	10.91	12.00
Non-Durable MFG	33	20.64	17.36
Wholesale Trade	42	5.07	3.82
Retail Trade	44	7.05	6.23
Transportation, Warehousing, Utilities	48	18.06	16.05
Information	51	15.52	9.97
Finance & Insurance	52	8.68	8.12
Real Estate	53	6.73	3.84
Pro Bus Services	54	14.13	10.28
Employment Services	56	53.12	25.77
Ed Services	61	55.68	41.08
Health Care	62	11.58	11.09
Arts, Entertainment, Recreation	71	45.49	40.49
Accommodation, Food Services	72	7.97	8.85
Other Services	81	6.32	6.20
Fed Government	91	17.61	15.89
State & Local Ed	92	136.99	113.54
State & Local Non-Ed	93	22.33	21.16

NOTE: No differences are statistically significant.

Reference-Attachment:

Developing a Birth/Death Model
Mark Crankshaw
BLS Washington
August 21, 2008

Background

Prior research has indicated that the current JOLTS estimation may not adequately capture the level of churning (hires and separations) actually occurring in the economy. This primarily due to the inability of the JOLTS survey to capture hires data from new and young firms and to capture separations from closing firms. Additionally, the divergence between the implied employment changes yielded through JOLTS hires and separations level estimates and the actual employment changes seen in CES estimation indicates that additional churning (primarily separations) is systematically under-reported to the JOLTS respondents. This finding was further confirmed by the recently conducted Response Analysis Survey (RAS) for the two industries with the largest divergence. These industries are Employment Services (ID56) and State and Local Government (ID92). While improvements in the JOLTS sampling methodology may help mitigate these inadequacies, the bulk of the shortcomings may have to be treated with a model.

To correct for the above inadequacies, a birth/death model has been developed that will address two separate shortcomings:

- The model will attempt to estimate for a given month the level of employment for firms entering the labor force (that is, birth employment). The model will also estimate the level of hires and separations for those birth establishments.
- The model will attempt to estimate for a given month the level of separations for firms exiting the labor force (that is, establishment deaths). Note that these establishments do not contribute to the employment level since firms that have exited the labor force have no employment by definition.

To that end, the LDB simulation of JOLTS hires and separations data will be utilized. (See the paper entitled ‘Simulating JOLTS Hires and Separations Data Using the LDB’ for the details of this method.) The simulation yields estimated employment, hires and separations for those establishments who have entered the labor force for a given month as well as the employment, hires and separations of those establishments that can not be adequately sampled (i.e., establishments less than 12 months old). The simulation also yields estimated separations levels for those establishments who have exited the labor force in a given month.

Birth Employment

The first aspect to be modeled is the level of birth employment (i.e., first time reporters as well as those young firms less than 30 months old) for a given industry for a given month. The birth employment level is taken directly from the monthly simulation of JOLTS data on the LDB.

Likewise, the hires and separations levels for the cohort of birth units were taken directly from the simulation.

Death Separations

The separations from the deaths on the LDB were drawn directly from the simulation. Only the first month of each quarter will contain deaths.

Forecasting

Since current LDB data is unavailable when JOLTS estimation is produced, it is not possible to simulate JOLTS birth/death employment, hires and separations. Therefore, it would be necessary to forecast JOLTS birth/death employment, hires and separations. One possible method that can be used to forecast this data would be to use an ARIMA prediction using historical JOLTS birth/death employment, hires and separations data. An ARIMA forecast has been conducted on this data and the forecast performed adequately. It is also possible to forecast using the ratio of CES year ago employment to current employment to adjust birth employment, hires, and separations.

Reference-Attachment:

Addressing JOLTS-CES Divergence

Beginning with the release of January 2009 data on March 10, BLS will implement improvements to the methodology used to generate estimates of hires, separations, and job openings from the Job Openings and Labor Turnover Survey (JOLTS). These changes are designed to improve the measurement of hires, separations, and openings and to more closely align the hires and separations estimates with monthly employment change as measured by the BLS establishment survey.

Research comparing the relationship between JOLTS hires and separations to the monthly employment change measured by the Bureau's Current Employment Statistics (CES) program (the establishment survey) indicate substantial discrepancies in employment trends over time. While JOLTS does not produce estimates of month-to-month change in employment, an implied employment change can be derived from JOLTS data by subtracting the separations estimate from the hires estimate for a given month. When viewed over time, this derived JOLTS measure of employment change does not track well with the CES, the Bureau's larger and better-known establishment survey. The CES is designed specifically to measure month-to-month employment change, collects data from a much larger sample, and benchmarks annually to universe employment counts, making CES the more reliable source of monthly employment change. Further, comparison of JOLTS hires and separations data to similar data produced in the Bureau's Current Population Survey (CPS or household survey) also indicates that JOLTS may be understating the levels of hires and separations.

BLS engaged in a multi-year research project to better understand these two issues, to establish their probable causes, and to develop improvements. As a result of this research, BLS plans to implement improvements in the following areas:

1) Revision of the JOLTS sample design to incorporate new business births more quickly, and to remove business deaths from the frame on a more timely basis;
2) Addition of a birth/death model for JOLTS to provide an estimate of hires, separations, and openings for births which are too new to be captured by the sample and for deaths which often do not get reported during monthly sample collection;
3) Modification to data collection, editing, and review procedures in specific industries where research has indicated a prevalence of particular response errors; and
4) Establishment of a monthly alignment procedure that takes the CES employment change estimates into consideration.

Improvements to the JOLTS Sample Design

Currently, the JOLTS sample is constructed from individual panels of sample units drawn on an annual basis. The full sample consists of one certainty panel made up of large units selected with virtual certainty based on their size, and 24 noncertainty panels. Each year a new set of panels is drawn from the Bureau's Longitudinal Database (LDB), a product of the Quarterly Census of Employment and Wages (QCEW) program. Each month a new noncertainty panel is rolled into collection, and the oldest noncertainty panel is rolled out. The collection life of a sample panel is therefore 24 months. This means that at any given time the JOLTS sample is constructed from panels from three different sampling frames, the most current being slightly over one year old and the oldest being slightly over three years old. Thus the JOLTS sample design reflects established firms that have been in business for a minimum of one year.

To better reflect the impact of younger establishments in the JOLTS sample, BLS is modifying the JOLTS sample design in the following ways. First, when a new set of panels is selected each year, the birth units in the sample (those not in existence on the previous year's frame) will be initiated for collection first, rather than waiting until their associated panel is initiated. Second, each quarter the newly updated LDB will be reviewed to identify birth establishments and a supplemental sample of these units will be drawn and added to the survey; at the same time, out-of-business units will be dropped from the sample on a quarterly basis. Thus, the JOLTS sample will be refreshed quarterly rather than annually. Third, the entire sample of old plus new panels will be poststratified and re-weighted annually to represent the most recent sampling frame; at present, this is not done for sample drawn from earlier frames. This procedure will make the sample more efficient than at present.

JOLTS Business Birth/Death Model

As with any sample survey, the JOLTS sample can only be as current as its sampling frame. The sampling frame for JOLTS is drawn from the LDB, which is updated quarterly from files submitted to the BLS QCEW program as part of the State Unemployment Insurance system. The built-in time lag from the birth of an establishment until its appearance on the sampling frame is approximately one year. In addition, many of these new units may fail within the first year. Since these universe units cannot be reflected on the sampling frame immediately, the JOLTS sample cannot capture job openings, hires, and separations from these units during their early existence. To develop data for these units that cannot be measured through sampling, BLS has developed a model to estimate the contribution of these units to the current month estimates. The birth/death model estimates birth/death activity for current month by examining the birth/death activity from previous years on the LDB and projecting forward using the ratio of over-the-year CES employment change. The birth/death model also uses historical JOLTS data to estimate the amount of "churn" (hires plus separations) that exists in establishments of various sizes. The model then combines the estimated churn with the projected employment change to estimate the number of hires and separations taking place in these units that cannot be measured through sampling.

The model-based estimate of total separations is distributed to the three components: quits, layoffs, and other separations, in proportion to their contribution to the sample-based estimate of total separations. Additionally, job openings for the modeled units are estimated by computing the ratio of openings to hires in the collected data and applying that ratio to the modeled hires.

The estimates of job openings, hires, and separations produced by the birth/death modeling process will then be added to the sample-based estimates produced from the survey to arrive at the final estimates for hires, separations, and openings.

Because JOLTS estimates did not previously include this step, addition of the birth/death model will raise the levels and rates of the hires, separations, and openings measured by JOLTS, and allow the series to more accurately reflect the current labor market.

Modifications to Data Collection Procedures

As stated earlier, an implied measure of employment change can be derived from the JOLTS data by subtracting separations from hires for a given month. Aggregating these monthly changes in the current series, however, generally produces employment levels that overstate employment change as measured by CES, at the total nonfarm level. Research into this problem has shown that a significant amount of the divergence between the CES employment levels and the derived JOLTS employment levels can be traced to the Employment Services industry and to the State Government Education industry. In the former industry, businesses have a difficult time reporting hires and separations of temporary help workers. In the latter industry, employers have a difficult time reporting hires and separations of student workers. BLS plans to devote additional resources to the collection, editing, and review of data for these industries. BLS analysts will more closely examine reported data that do not provide a consistent picture over time, and will re-contact the respondents as necessary. Analysts will work with the respondents to adjust their reporting practices as possible. Units that cannot be reconciled but are clearly incorrect on a consistent basis will be dropped from the estimation process and imputed for using existing techniques.

Establishment of an Alignment Procedure

Over time, employment change derived from JOLTS hires minus separations should track well with employment change measured through the CES. However, there are some definitional differences between the series that can cause legitimate differences for individual months. The major reasons for these month-to-month divergences are:

1) The reference periods of the two surveys are different. CES measures employment for the pay period including the 12th of the month, while JOLTS measures hires and separations for the entire month.
2) CES counts those who worked or received pay for the reference pay period, while JOLTS counts those who were hired or separated during the reference month. It is possible for a person to miss being paid for a given pay period without having been separated.

Both of these definitional differences can result in differing seasonal patterns between the two series, and therefore cause JOLTS to diverge from the CES in the short-term. Over time however, the computation of JOLTS hires minus separations should reflect employment changes that are consistent with the trends measured by the CES. The three changes to JOLTS that have been described above are expected to produce JOLTS series' that are much more consistent with the CES. The residual divergence will be controlled through a monthly alignment procedure that allows JOLTS to vary from CES for the reasons listed above, while ensuring that the long-term trends in JOLTS hires-minus-separations match those of the CES net employment change.

The goal of this process is to use current monthly CES employment trends to align the JOLTS implied employment trend (hires minus separations) to be approximately the same, but without forcing all the seasonal patterns to be the same between the surveys. This method takes advantage of the fact that the CES employment series for the current reference month is available prior to the production of JOLTS estimates for that same reference month.
The method works as follows:

- Each month, the initially computed seasonally adjusted JOLTS hires-minus-separations employment change estimate is adjusted to equal the CES seasonally adjusted net employment change estimate, through a proportional adjustment of the hires and separations estimates. By comparing the JOLTS and CES seasonally adjusted changes, the alignment procedure preserves legitimate differences in the seasonal patterns of underlying JOLTS and CES
- Proportional adjustment means that the two components (hires, separations) are adjusted in proportion to their contribution to the total churn (hires plus separations). For example, if hires is 40% of the churn for a given month, it will receive 40% of the needed adjustment and separations will receive 60% of the needed adjustment.
- In the next step, these adjusted hires and separations estimates are converted back to not seasonally adjusted data by reversing the application of the original seasonal adjustment factors.
- These trend-corrected not seasonally adjusted series are then put through the standard X-12 ARIMA seasonal adjustment process to create the final seasonally adjusted published series. These final seasonally adjusted series will not precisely equal the CES seasonally adjusted net employment change but will be very similar.

Revisions to Historical Series

The monthly JOLTS series begin with estimates for December 2000. All published estimates back to that point will be revised to reflect the addition of the revised birth-death model and the new alignment procedure, as well as selected adjustments to individual survey reports. New historical series for job openings, hires, total separations, quits, layoffs and discharges, and other separations will replace the currently available series. At that time, tables comparing the original and revised series will also be available.

Reference-Attachment:

Proposed JOLTS Sample Weights Adjustment
Sarah Goodale
July 2008

Background:

The Job Opening and Labor Turnover Survey (JOLTS) is a stratified random sample with a sample size of 16,000 establishments. The 16,000 establishments are distributed over 25 panels; in which 1 panel is a certainty panel and the remaining 24 panels are noncertainty panels. Each month one panel enters the sample (rolls in) while another panel leaves the sample (rolls out).

Each year a sample is sample is drawn, with which 12 panels will be used to enter the JOLTS sample. Since there are 24 panels that are in rotation, 12 panels of the sample can come from the new sample while the remaining panels are from previous samples. There is a possibility that there are 3 different samples present in JOLTS at once. When the first month of the new sample rolls into to JOLTS; there is 1 panel of the new sample, 12 panels of the sample taken the previous year, and 11 panels of the sample taken 2 years prior. Since the sample weights for JOLTS is currently determined when the establishments are selected to be a part of JOLTS, there can also be three different frames in which the establishments weight to. Also once an establishment has been rolled into JOLTS; it is only removed when the panel rolls out of JOLTS.

Younger establishments are represented proportionally for the frame on the current yearly sample selected. However, when this sample is added to the older samples to make up the 24 panels of JOLTS, the younger establishments are then disproportionate to the frame. Also the younger establishments are mostly represented in JOLTS by the most current sample and are not distributed among the different panels of JOLTS. The younger establishments may have different characteristics then the older establishments, and therefore should be properly represented on the sample.

Objectives:

1) To weight all establishments in JOLTS to the current frame
2) To weight the younger establishment to the represent the appropriate amount on the current frame for all 24 panels
3) To provide a birth refresh of new establishments to help improve the distribution of younger units

Procedure:

1) Draw the new annual sample
    a) Draw the sample using the current sampling procedure
    b) Keep the full frame file
    c) Keep the full 24 panel sample
2) Update the previous samples

  a) Create a subset containing the previous two samples
  b) Remove any Out-of-Business Establishments
  c) Place the establishments in the proper stratum
    i) Merge the previous sample subset with the current full frame, keeping the stratum definition of the current frame dataset

3) Assign the age variable
  a) Assign the age variable to each of the datasets
    i) Age = 0 : establishments that come into existence on the JOLTS frame for the first time or since the last frame to the current frame
    ii) Age = 1: establishments that have been on the JOLTS frame for a year
    iii) Age = 2: establishments that have been on the JOLTS frame for at least two years
  b) Assign a post stratification variable to the samples and the frame
    i) Age = 0 or Age = 1 post stratification is age/industry/size
    ii) Age = 2 post stratification is age/region/industry/size

4) Assign the panel to the new 12 panel sample (old samples only have the weights appended)
  a) Separate the new sample (24 panels drawn earlier) remove the certainty units from the sample and find the count of establishments per stratum
    i) Divide the count of establishments by 24 call this amt
  b) Separate the new sample into the groups age = 0 and 1 and age = 2
    i) If age = 0 or 1 then keep only the first 12 panels
  c) All Age = 0 go into panel 1
    i) Assign a new schedule number to the establishments
  d) All Age = 1 go into panel 2
    i) Assign a new schedule number to the establishments
  e) All Age = 2 go into panel 3 – panel 12
    i) Create amount = 10*amt (amt is the number per stratum in each panel)
    ii) Sort the data in age = 2 into the post stratification variable in order of there original schedule number
    iii) Assign a sequence number to these elements in the post stratification variable.
      (1) Keep only those elements whose sequence number is less than or equal to the amount
    iv) Assign the elements to new panel numbers
    v) Join the sample with the certainty units, and age = 0 and 1.
    vi) Assign a new schedule number to elements in the sample

5) Calculate the new weight
  a) Join the new sample with the previous sample
  b) Find the counts of the post stratification variable for the frame and the sample (panel 3 – 26 of the 36 panels of the 3 samples)
  c) Using the post stratification calculate the new weights

$$weight_{age_{post_{strata}}} = \frac{N_{post_{strata}}}{n_{post_{strata}}}$$

6) Birth refresh: This will be done in between the yearly samples
  a) Pull all units from the quarter of interest (from the LDB)
  b) Assign the age variable, keeping only the units that are Age = 0
  c) Remove any OOB and OOS units

d) Assign the post stratification units to the units
e) Find the counts for the post stratification variable
f) Find the amount to sample of birth per post strata by

$$n_{birth_{post_{strata}}} = \frac{N_{Q, post_{strata}}}{weight_{age_{post_{strata}}}}$$

g) Distribute the births in the 3 panels
   i) Q2 birth panel 4 – panel 6
   ii) Q3 birth panel 7 – panel 9
   iii) Q4 birth panel 10 – panel 12

7) Create the new full sample file

Reference-Attachment:

JOLTS Item Imputation
Alternative Approach
Mark Crankshaw
Summer-Fall 2015

Introduction

The Job Openings and Labor Turnover Survey (JOLTS) is produced by the US Bureau of Labor Statistics (BLS). Each month, since the inception of the JOLTS program in 2000, BLS has collected job openings, hires and separations data through a sample of approximately 16,000 nonfarm business establishments. The job openings data help assess the demand for labor in the US labor market while hires and separations data help measure the churn in the labor force.

Item non-response in JOLTS is dealt with primarily through a hot deck nearest neighbor algorithm whereby each missing value is replaced with an observed response from a “similar” unit. Over the course of the survey, approximately 12% of JOLTS item data has been imputed.

Why have an alternative?

Item imputation donors and recipients are paired together in JOLTS based on ***reported employment*** within an industry (NAICS super-sector level) imputation cell. The drawback of this approach, within the JOLTS context, is that the ***level*** of reported employment is not a useful measure for what is of primary interest in JOLTS, namely, measuring the effects of ***employment dynamics***. By employment dynamics, what is meant is whether an establishment is stable, growing, or is contracting as a function of over-the-month reported employment change. Approximately 36% of JOLTS responding establishments report no over-the-month employment change while the remainder of respondents are evenly split between those that report positive over-the-month employment growth and negative over-the-month employment growth.

In theory, JOLTS variables of interest such as job openings, hires and separations should be correlative to the employment dynamics of a given establishment. Stable establishments should tend, as a group, to report hires rates that are approximately equal to their reported separations rates. Expanding establishments should tend, as a group, to report hires rates in excess of their reported separations rates. Likewise, contracting establishments should tend, as a group, to report separations rates in excess of their reported hires rates.

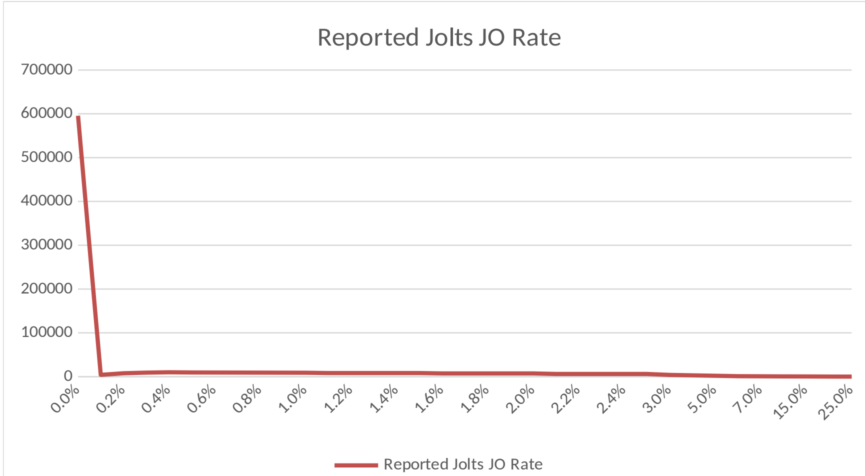
Empirically, JOLTS historical microdata does, in fact, strongly and consistently align with the theoretical relationships outlined above. Stable establishments tend to have relatively low levels of job openings, hires and separations and, in aggregate, stable establishments tend to have reported hires rates approximately equal to the reported separations rates. Consistent with the

theoretical framework, expanding establishments tend to have as a group reported hires rates significantly higher than reported separations rates and contracting establishments as a group tend to report higher separations rates than hires rates.
The current nearest neighbor imputation algorithm, however, is oblivious towards the employment dynamics of both donors and recipients. Although two establishments may share a common level of reported employment, this in no way implies that they share a common type of employment dynamics. It is as likely that two establishments with a common level of reported employment have diametrically opposed employment dynamics as have similar employment dynamics. The current nearest neighbor imputation approach relies solely on reported employment, which carries little JOLTS informative value, and completely disregards employment dynamics, the very variable that encapsulates considerable, perhaps even the most, JOLTS informative value.

The item that is “borrowed” from the donor by nearest neighbor recipients in JOLTS is the rate of each item (where rate= $\frac{item}{reported\ employment}$ ¿. The distribution of rates for JOLTS items is non-Normal. (See the graph entitled ‘Reported JOLTS JO Rate’ below for a typical example.) Over 60% of item responses are reported as zero, indicating that for that month the establishment had no job openings (or hires, quits, etc.). The remainder of the distribution is:

1. Nearly uniformly distributed,
2. Multimodal with modal points of extremely low amplitude (for example, where one hire is reported, there are many modes in reported hires rate at $\frac{1}{n}$, where employment n goes from 1,…,∞),
3. Asymmetrical and severely skewed right.

The frequency of observations associated with a given rate steadily falls as the rate increases. However, the length of the right-hand tail is very long. The graph below is a truncated example of the distribution as JOLTS rates of 50-200% are not uncommon, particularly with establishments that report few employees. In rare cases, the reported rate can legitimately exceed 1000% - 8000%. The range of reported rates varies significantly by size. Smaller establishments have more variation in reported rates, since if employment is at low level (1-5 employees), any amount of JOLTS activity (1 or more reported job openings, hires, etc.) results in reported rates that are multiples of the cell mean rate. Very high rates are atypical in larger establishments.



The extreme length of the right-hand tail of the JOLTS variable distributions is responsible for another drawback of the Nearest Neighbor imputation algorithm. Establishments can be paired that have diametrically opposed employment dynamics. Additionally, donor rates can be borrowed very far into the extreme right-hand tail of the distribution. The JOLTS Program Office has noted that this borrowing from the extreme right-hand tail can produce unexpected volatility in JOLTS estimates and, further, that treatment of the resultant unexpected volatility is time consuming and a strain on production resources.

From a theoretical perspective, the probability of a donor reporting a rate from the extreme right of the rate distribution serves to compound the theoretical shortcoming alluded to earlier. An extremely high level of hires, for example, should be correlated with high levels of employment growth, while extremely high levels of separations, for example, should be correlated with high levels of employment decline. Since there is a probability that a donor has a rate in the extreme right-hand tail, it is very unlikely that donor will be matched with a recipient with the appropriate level of employment dynamics (in terms of both direction AND magnitude). Stated another way, it is very likely in the case that a donor reporting extremely high rates will be paired up with a recipient with a completely inappropriate level of employment dynamics (in both direction and magnitude).

In the “ideal” JOLTS nearest neighbor imputation approach, donors and recipients would be paired based on similar employment dynamics. We would ideally like to have contracting donors paired with contracting recipients, expanding donors paired with expanding recipients, and stable donors paired with stable recipients. However, the current imputation cell in JOLTS is the Industry cell and even at that very high level of aggregation there is a low ratio of donors to recipients in some industries. Dividing the Industry further into three dynamics groups would reduce the ratio of donors to recipients in all imputation cells and this reduction in the donor pool would increase the probability of extreme right tail donors being selected as nearest neighbor donors. While the “ideal” approach would positively impact the theoretical drawbacks of the current nearest neighbor imputation approach, it would negatively impact the volatility in JOLTS estimates by increasing the probability of extreme right tail donors being selected as nearest neighbors and increase of the probability of inappropriately pairing establishments of differing sizes. Restricting the donor pool by size also would have the same negative effects.

Proposed Alternative

The proposal offered in this paper is to replace the hot deck nearest neighbor imputation approach currently being used in JOLTS with a simple respondent data driven model. It is hoped that this model can retain all of the positive aspects of the nearest neighbor approach while correcting the drawbacks identified in the previous section. Specifically, the model: 1) should be based on currently reported data enabling the model to capture changes in reporting patterns as they occur, 2) should not adversely affect variance estimation, 3) should allow recipients to receive imputed data from donors with similar employment dynamics, and 4) should be able to lessen the impact of donors drawn from the extreme right-hand tail of the rate distribution. This model is not intended to be the "final solution" to all JOLTS imputation short-comings but rather should be seen as an informal bridge to more sophisticated formalized statistical models. Until those can be developed and implemented, however, the model being proposed here can improve JOLTS imputation with respect to theory and actual practice; it is an attempt to improve the current JOLTS imputation results.
At the heart of this simple model is the concept of producing three separate models for each of the JOLTS industry imputation cells. One model based on the respondent rate distribution of stable respondents, another based on the respondent rate distribution of expanding respondents, and the third based on the respondent rate distribution of contracting respondents with the employment dynamics classification based on the reported over-the-month employment change of the respondents.

Suppose that $\theta_{id,t}$ represents a variable of interest in JOLTS (job openings, hires, etc.) in industry *id* for a given month *t.* JOLTS item imputation is concerned only with those sampled establishments that reported *at least* employment. Complete (or unit) non-respondents are accounted for in JOLTS using a non-response adjustment factor (NRAF). Therefore, for each variable$\theta_{id,t}$, respondent establishments can be classified as either item respondents ¿¿) or as item non-respondents ¿¿). Suppose that $\varepsilon_{id,t}$ represents reported employment for a JOLTS respondent within a given industry *id* and given month *t* and that $\varepsilon_{id,t-1}$ represents reported employment for the same JOLTS respondent within a given industry *id* and in the previous month. We can then define employment change as: $(\varepsilon$¿¿$id,t-\varepsilon$¿¿$id,t-1)$¿¿.

The current nearest neighbor JOLTS algorithm pairs a member of*the set of* ¿¿) with a member of the set of ¿¿) based on the "nearness" of the reported value of $\varepsilon_{id,t}$ for each pair. The item imputation donor "donates" the value of $\frac{\theta_{id,t,ir}}{\varepsilon_{id,t}}$ to the item imputation recipient. We will denote the donor rate¿)*100 asω. The item imputation recipient thus "borrows" ω and the imputed level (Ω) is found by multiplying the borrowed donor rate (ω) by the recipient's reported employment:

Imputation level (Ω) = $\varepsilon_{id,t,inr} * \omega$

The proposed alternative to the current imputation approach is to subdivide the current industry imputation cell into three parts based on the reported employment change $(\varepsilon¿¿id,t-\varepsilon¿¿id,t-1)¿¿$ for each respondent establishment.

Where: $\varepsilon_{id,t} - \varepsilon_{id,t-1} = 0$ is the stable group with donor rates denoted as $\omega_{i,id,stbl}$
$\varepsilon_{id,t} - \varepsilon_{id,t-1} > 0$ is the expanding group with donor rates denoted as $\omega_{i,id,\exp}$
$\varepsilon_{id,t} - \varepsilon_{id,t-1} < 0$ is the contracting group with donor rates denoted as $\omega_{i,id,con}$

where i=1,…,n given n donor establishments within the id/group.

A simple model of the item respondent rate distribution will then be constructed based on the known distribution characteristics of the item respondent rates for each group ($\omega_{i,id,stbl}, \omega_{i,id,\exp} \wedge \omega_{i,id,con}$). Independent draws from the model distribution will then be applied to the imputation recipient employment to produce imputed levels. The variables of interest within the actual reported donor data used to construct the model distribution are:

1) The mean rate of the items within each group ($\mu_{\omega,id,group}$) which is calculated as sum of the item over all item respondents divided by the employment of all item respondents, $\frac{\sum \theta_{id,group,ir,t}}{\sum \varepsilon_{id,group,ir,t}}$
2) The standard deviation of the absolute differences between item respondent rate and the cell mean within each group ($\sigma_{\omega,id,group}$) which is calculated as follows:

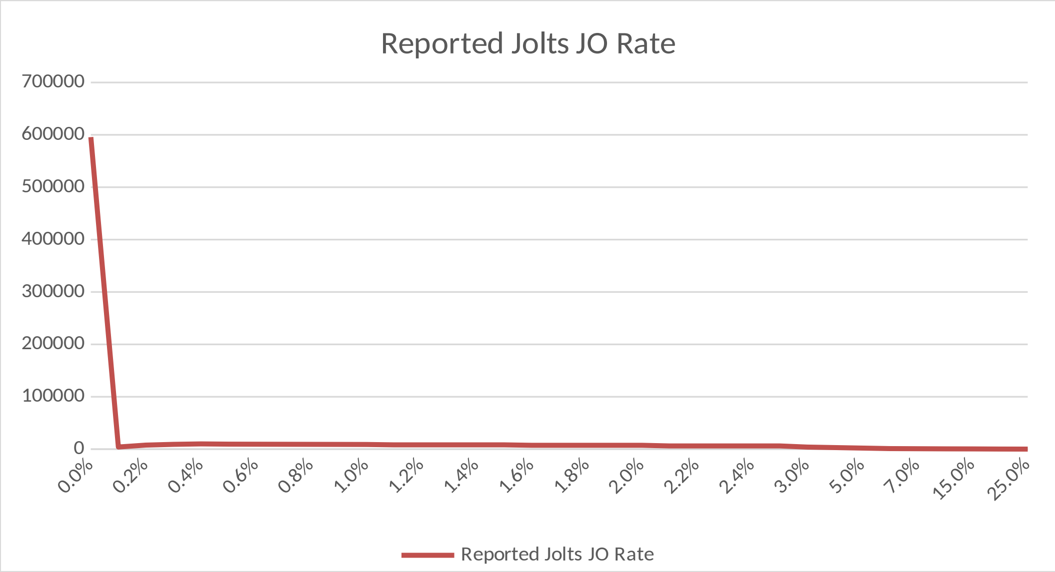
ABS ($\frac{\sum \theta_{id,group,ir,t}}{\sum \varepsilon_{id,group,ir,t}}$ - $\frac{\theta_{id,group,ir,t}}{\varepsilon_{id,group,ir,t}}$), for each *ir* in each *id,group,ir* cell.

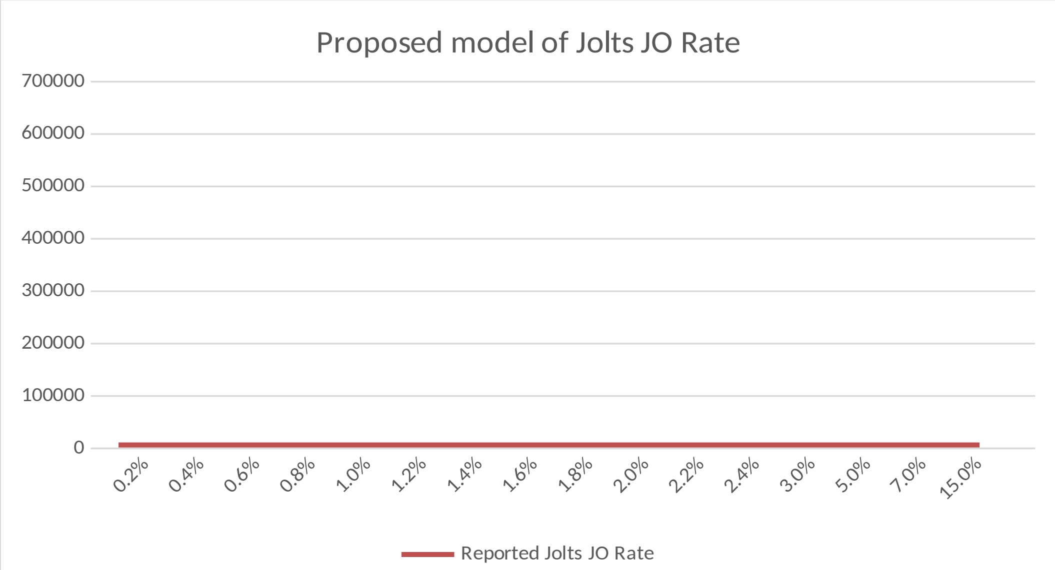
3) The percentage of item respondents who report rates below the cell mean within each group is represented by ($\beta_{id,group}$).
4) The percentage of item respondents who report rates above the cell mean within each group is represented by ($1-\beta_{id,group}$).

As noted earlier, we wish to construct a model that does not adversely affect variance estimation and we wish to construct a model that lessens the unexpected volatility created by donors drawn from the extreme right-hand tail of the rate distribution. A compromise between these two conflicting goals can be accommodated by slightly altering the model distribution from the actual item respondent rate distribution. Response Analysis Survey (RAS) studies have indicated that item respondents to JOLTS may systematically differ from item non-responders with respect to reporting ‘0’ values to JOLTS. Reporting a value of ‘0’ to JOLTS is an indication that the establishment has had no JOLTS activity (no job openings have been made available, no hires were made, etc.). For most item respondents, particularly with smaller and stable establishments, there is little effort needed on the part of respondents to collect the data necessary to report a lack of JOLTS activity. The RAS studies, however, have indicated that establishments that report missing values are unlikely to have a lack of JOLTS activity, rather that these establishments have JOLTS activity (job opening, hires, etc.) and that these respondents are unwilling or unable to quantify this activity as a JOLTS item response. While about 60% of JOLTS item responses are ‘0’ there is evidence that the propensity to report ‘0’ is far less for item non-respondents.

The compromise proposed in the model distribution is to exclude ‘0’ responses from the model in exchange for truncating a portion of the extreme right-hand side of the rate distribution (the source of the unexpected volatility). In the Balanced Half Sample replication variance estimation technique employed by JOLTS, multiple reports of ‘0’ will tend to depress variance estimates. Removing ‘0’ from the donor rate distribution will have the effect of increasing variance estimates. Truncating a portion of the extreme right-hand side of the rate distribution will have the opposite effect; it will tend to depress variance. Therefore, removing ‘0’ from the model donor rate distribution while truncating a portion of the extreme right-hand side of the donor rate distribution produces offsetting effects. Given that imputed item variables represent a small portion (12%) of overall JOLTS data, a reasonable expectation is that removal of the ‘0’ from the model while truncating the extreme right tail will not unduly affect the overall JOLTS variance estimates.

To simplify the proposed model further, we will explicitly create a model based on a uniform distribution by removing the myriad of low amplitude modal points exhibited in the reported data. The graphs below detail the nature of the actual reported job openings rate versus the proposed model of the JOLTS:

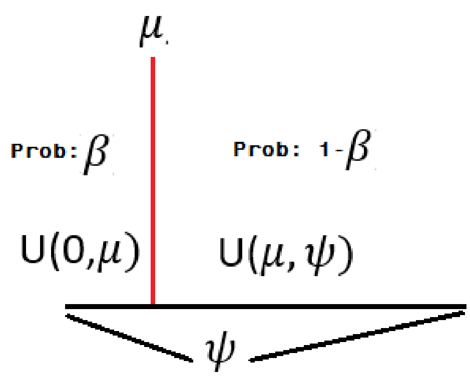




The length of the proposed model distribution ($\psi_{id,group}$) is dependent upon the standard deviation of the reported donor data within the industry-dynamic group imputation cell ($\sigma_{\omega,id,group}$). The length of the model distribution is determined by the 90% confidence interval of ψ around the id/group cell mean¿). The uniform distribution in the model will be fit, with the lower bound of the uniform at 0 and the upper bound at the length, to form the uniform distribution U(0,$\psi_{id,group}$). The effect of this model distribution is to truncate a portion of the right-hand tail of the actual donor rate distribution within each id/group cell.

Taking into account the percentage of item respondents within a group which report rates below the cell mean ¿¿) enables the model uniform distribution U(0,ψ) to be broken into two parts and helps the model distributions to more closely conform to the actual reported rate distribution of item respondents with respect to the skewness of the distribution. For each dynamic group within an industry, we can calculate the cell mean and the percentage of item respondents that have reported above and below the cell mean. We can then utilize that information to construct two model uniform distributions for each id/group imputation cell: one from the zero bound to the cell mean U(0,μ], and the other from the cell mean to the length of the overall uniform distribution U(μ , ψ).

The model is then constructed such that the probability that imputed values drawn from U(0,μ] is the same as the percentage (β) of actual item reporters that report rates from 0 to the cell mean and the probability that imputed values drawn from U(μ , ψ) is the same as the percentage (1-β) of actual item reporters with reported rates above the cell mean (as illustrated in the figure below). This process allows for the pattern of skewness in the item respondent rates to be approximately replicated in the imputed rates that will be drawn from the model.



The final consideration in the model is an accounting for differences in reported item rates with respect to establishment size (where size class is determined by reported employment). As mentioned earlier, JOLTS item rates vary by size of establishment, with smaller establishments reporting greater variation in rates. We can classify each respondent into a size class using the same size criteria used in JOLTS sampling:

- $\varepsilon_{id,t,r}$ in (0,9) classified into size class 1
- $\varepsilon_{id,t,r}$ in (10,49) classified into size class 2
- $\varepsilon_{id,t,r}$ in (50,249)classified into size class 3
- $\varepsilon_{id,t,r}$ in (250,999) classified into size class 4
- $\varepsilon_{id,t,r}$ in (1000,4999)classified into size class 5
- $\varepsilon_{id,t,r}$ >= 5000 classified into size class 6

Once each reporter is classified by establishment size, we then pool all item respondents across industry and dynamic groups, and calculate the standard deviation of item rates by size ($\sigma_{\omega,¿¿¿}$). We can then calculate the length of the six size class distributions ($\psi_{¿¿.¿}$ Next, we compare each of the id/group lengths $\psi_{id,group}$ for each item respondent within each size class. If the length of the id/group for that establishment is greater than the length for the size class, then the length for the id/group is made equal to the length of the size class distribution:

$$\text{If} \quad \psi_{id,group} \; > \; \psi_{¿¿¿} \quad \text{then} \quad \psi_{id,group} = \psi_{¿¿¿}$$

The smaller establishments typically have larger values for $\psi_{¿¿¿}$ (since their rate distribution typically has higher variance), so the net effect of this adjustment is to keep the length of the distributions for smaller establishments intact while lowering the $\psi_{id,group}$ values for the larger establishments (since the rate distributions of larger establishments typically have less variance) to match the observed differences in the actual rate distributions between smaller and larger establishments.

Example

I will illustrate the proposed alternative using actual JOLTS data. The attached spreadsheet contains microdata for a randomly chosen industry for a randomly chosen month (ID 44 Retail Trade, April 2011). The spreadsheet contains the microdata for item respondents and item non-respondents, and the calculations described here can be found in the spreadsheet.

We begin by sorting item respondents and item non-respondents into their proper group based on their over-the-month employment change$\left(\varepsilon ¿¿ id,t-\varepsilon ¿¿ id,t-1\right) ¿¿$. Each tab in the spreadsheet contains the microdata for the group (tab 'Group C' contains the microdata for contracting establishments where $\varepsilon_{id,t}-\varepsilon_{id,t-1}$ < 0, tab 'Group E' contains the microdata for expanding establishments where $\varepsilon_{id,t}-\varepsilon_{id,t-1}$ > 0, and tab 'Group S' contains the microdata for stable establishments where $\varepsilon_{id,t}-\varepsilon_{id,t-1}$ = 0). There is a fourth tab, 'ALL DATA', containing the microdata for all of the establishments within the industry, including those establishments that have no employment change (since they have no previous month's employment).

Next, we can summarize the set of item respondents with respect to the mean rate of the items within each group ($\mu_{\omega,id,group}$), the standard deviation of the absolute differences between item respondent rate and the cell mean within each group ($\sigma_{\omega,id,group}$), and the percentage of item respondents who report rates below the cell mean within each group ($\beta_{\omega,id,group}$).

GROUP: CONTRACTING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	n	μ	σ	β
Job Openings	133	3.99%	5.43%	76%
Hires	198	3.59%	5.61%	70%
Quits	174	2.40%	3.98%	59%
Layoffs & Discharges	17	1.34	31.29	80%

	4	%	%	
Other Separations	170	0.62%	49.76%	92%

GROUP: EXPANDING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	n	**μ**	**σ**	**β**
Job Openings	173	4.70%	3.05%	85%
Hires	257	4.43%	13.56%	50%
Quits	218	1.61%	2.07%	64%
Layoffs & Discharges	218	1.01%	1.21%	85%
Other Separations	212	0.07%	0.59%	94%

GROUP: STABLE ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	n	**μ**	**σ**	**β**
Job Openings	327	1.22%	3.05%	94%
Hires	351	2.05%	4.08%	85%
Quits	342	1.42%	3.54%	85%
Layoffs & Discharges	342	0.69%	5.40%	93%
Other Separations	340	0.14%	0.36%	97%

This industry is typical of JOLTS data with respect to the relationship between the cell means of hires and separations (the sum of quits, layoffs, and other separations). In the contracting group, the mean hire rate (3.59%) is less than the sum of the separations components means (4.36%). The expanding group has a mean hires rate (4.43%) greater than the sum of the separations components means (2.68%) while the mean hires rate of stable establishments (2.05%) is approximately equal to the sum of the separations component means (2.25%).

The JOLTS variable total separations represents the summation of its components of quits, layoffs and discharges, and other separations. In the current nearest neighbor approach, the components of total separations are imputed while total separations are not. This proposed alternative approach does not impute for total separations either. Instead, each component of total separations is independently imputed. If total separations is reported and any component of separations is not, then the imputed components levels will then be scaled so that their sum equals the reported total separations level as is done in the current imputation approach.

Next, we can construct the length of the proposed model distribution ($\psi_{id,group}$) using the observed deviations (**σ**). The length (**ψ**) is equal to the value 1.645*(**σ**) for a 90% confidence interval.

GROUP: CONTRACTING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	n	μ	σ	ψ	β
Job Openings	133	3.99%	5.43%	8.93%	76%
Hires	198	3.59%	5.61%	9.23%	70%
Quits	174	2.40%	3.98%	8.37%	59%
Layoffs & Discharges	174	1.34%	31.29%	51.89%	80%
Other Separations	170	0.62%	49.76%	81.86%	92%

GROUP: EXPANDING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	n	μ	σ	ψ	β
Job Openings	173	4.70%	3.05%	5.01%	85%
Hires	257	4.43%	13.56%	22.30%	50%
Quits	218	1.61%	2.07%	3.41%	64%
Layoffs & Discharges	218	1.01%	1.21%	1.99%	85%
Other Separations	212	0.07%	0.59%	0.97%	94%

GROUP: STABLE ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	n	μ	σ	ψ	β
Job Openings	327	1.22%	3.05%	5.02%	94%
Hires	351	2.05%	4.08%	6.71%	85%
Quits	342	1.42%	3.54%	5.83%	85%
Layoffs & Discharges	342	0.69%	5.40%	8.88%	93%
Other Separations	340	0.14%	0.36%	0.97%	97%

Now we can construct the two Uniform distributions from which imputed rates can be drawn. With probability 1-β (the probability that an item respondent within the group has reported

above the cell mean) we will draw from U(0,μ), and with probability β (the probability that an item respondent within the group has reported **below** the cell mean) we will draw from U(μ, ψ).

GROUP: CONTRACTING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	μ	ψ	β	BELOW	1-β	ABOVE
Job Openings	3.99%	8.93%	76%	U(0,3.99%)	24%	U(3.99%,8.93%)
Hires	3.59%	9.23%	70%	U(0,3.59%)	30%	U(3.59%,9.23%)
Quits	2.40%	8.37%	59%	U(0,3.98%)	41%	U(3.98%,8.37%)
Layoffs & Discharges	1.34%	51.89%	80%	U(0,1.34%)	20%	U(1.34%,51.89%)
Other Separations	0.62%	81.86%	92%	U(0,0.62%)	8%	U(0.62%,81.86%)

GROUP: EXPANDING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	μ	ψ	β	BELOW	1-β	ABOVE
Job Openings	4.70%	5.01%	85%	U(0,4.70%)	15%	U(4.70%,5.01%)
Hires	4.43%	22.30%	50%	U(0,4.43%)	50%	U(4.43%,22.30%)
Quits	1.61%	3.41%	64%	U(0,1.61%)	36%	U(1.61%,3.41%)
Layoffs & Discharges	1.01%	1.99%	85%	U(0,1.01%)	15%	U(1.01%,1.99%)
Other Seps	0.07%	0.97%	94%	U(0,0.07%)	6%	U(0.07%,0.97%)

GROUP: STABLE ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	μ	ψ	β	BELOW	1-β	ABOVE
Job Openings	1.22%	5.02%	94%	U(0, 1.22%)	6%	U(1.22%, 5.02%)
Hires	2.05%	6.71%	85%	U(0, 2.05%)	15%	U(2.05%, 8.16%)
Quits	1.42%	5.83%	85%	U(0, 1.42%)	15%	U(1.42%, 5.83%)
Layoffs & Discharges	0.69%	8.88%	93%	U(0, 0.69%)	7%	U(0.69%, 8.88%)
Other Seps	0.14%	0.59%	97%	U(0, 0.14%)	3%	U(0.14%, 0.59%)

The final data required to produce imputed rates are the lengths of the six size class distributions (ψ¿¿¿ for each variable. These lengths are calculated the same way as the group lengths, however, these lengths are based on the data across all industries and groups for a given month. The matrix below details the size class lengths (size x variable matrix) for April 2011 in Industry A:

ψ

SIZE	JO	HIRES	QUITS	LD	OS
1	72.60%	28.11%	10.42%	28.52%	61.01%
2	38.07%	27.40%	5.90%	26.05%	2.58%
3	24.15%	11.45%	3.04%	10.50%	0.56%
4	15.54%	9.38%	2.30%	21.17%	0.55%
5	5.01%	7.11%	2.54%	4.41%	0.39%
6	2.43%	1.63%	0.72%	1.27%	0.37%

Now we can begin imputing rates. In the attached spreadsheet under tab ‘Group C’, the first schedule (20000949) listed is an item non-respondent for Job Openings. The non-respondent has reported 2,648 employees (classifying it as size class 5), and its over-the-month employment change is -3. It is therefore classified as a contracting establishment. The truncated table below will then be used to impute Job Openings for this non-respondent.

GROUP: CONTRACTING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	μ	**ψ**	**β**	**BELOW**	**1-β**	**ABOVE**
Job Openings	3.99%	8.93%	76%	U(0,3.99%)	24%	U(3.99%,8.93%)

In the ψ matrix above, we note that the Job Openings (JO) length (ψ) for size class 5 is 5.01%. Since the value ofψ for the ID 44/contracting group (8.93%) is greater than the size class 5 JO length (5.01%), we replace the upper bound value for the id/group with the ψ value of the size. The table below illustrates this adjustment:

GROUP: CONTRACTING ESTABLISHMENTS
ID: 44 RETAIL TRADE
SIZE CLASS: 5
MONTH: April 2011

	μ	**ψ**	**β**	**BELOW**	**1-β**	**ABOVE**
Job Openings	3.99%	5.01%	76%	U(0,3.99%)	24%	U(3.99%,**5.01%)**

We begin item imputation by drawing from a Uniform distribution, U(0,1), to determine if the imputed rate is to be drawn from the “BELOW” uniform distribution or the “ABOVE” uniform distribution. If the value of the U(0,1) is less than or equal to β then we will draw from the “BELOW” uniform distribution, otherwise we will draw from the “ABOVE” distribution.

If U(0,1) > β then draw from U(3.99%,5.01%)
If U(0,1) <= β then draw from U(0,3.99%]

Suppose that the value drawn from U(0,1) is 0.6616, then since 66% is less than β (76%), we will draw from the “BELOW” distribution, U(0,3.99%]. Suppose that the value drawn from U(0,3.99%] is 2.854%. The imputed **rate** for Job Openings for schedule 20000949 is thus 2.854%, while the imputed level is the reported employment of the recipient establishment multiplied by the imputed rate, so the imputed Job Openings **level** is 2648*2.854%=75.57.

As a second example, consider schedule 70790621 (found in the Excel spreadsheet, tab ‘Group E’, 9th establishment in the list). This establishment is an item non-respondent for Quits. The non-respondent has reported 111 employees (classifying it as size class 3), and its over-the-month employment change is +1. It is therefore classified as an expanding establishment. The truncated table below will then be used to impute Quits for this non-respondent.

GROUP: EXPANDING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	μ	**ψ**	**β**	**BELOW**	**1-β**	**ABOVE**
Quits	1.61%	3.41%	64%	U(0,1.61%)	36%	U(1.61%,3.41%)

In the ψ matrix, we note that the Quits (Q) length (ψ) for size class 3 is 3.04%. Since the value of ψ for the ID 44/expanding group (3.41%) is greater than the size class 3 Quits length (3.04%), we replace the upper bound value for the id/group with the ψ value of that size. The table below illustrates this adjustment:

GROUP: EXPANDING ESTABLISHMENTS
ID: 44 RETAIL TRADE
MONTH: April 2011

	μ	**ψ**	**β**	**BELOW**	**1-β**	**ABOVE**
Quits	1.61%	3.04%	64%	U(0,1.61%)	36%	U(1.61%,**3.04%**)

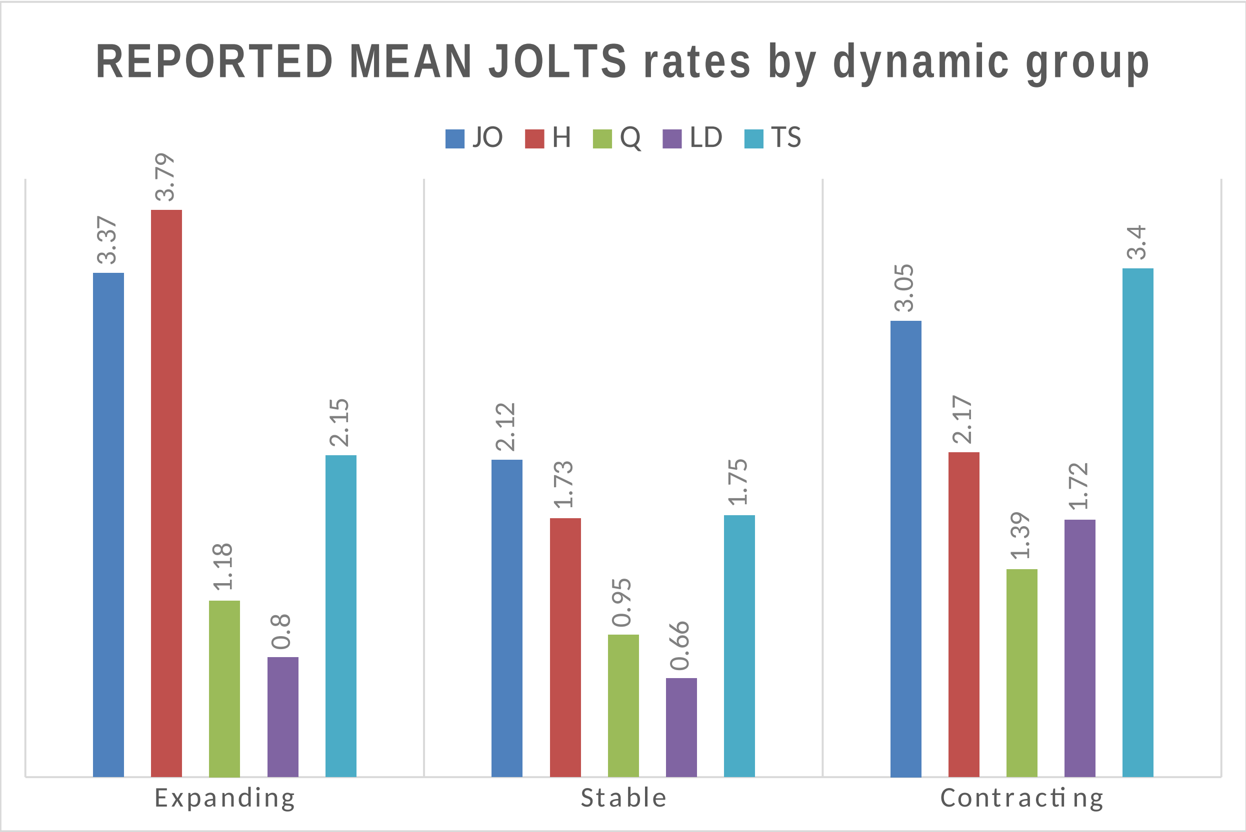
As before, we begin item imputation by drawing from a Uniform distribution, U(0,1), to determine if the imputed rate is to be drawn from the “BELOW” uniform distribution or the “ABOVE” uniform distribution. If the value of the U(0,1) is less than or equal to β then we will draw from the “BELOW” uniform distribution, otherwise we will draw from the “ABOVE” distribution.

If U(0,1) > β then draw from U(1.61%,3.04%)
If U(0,1) <= β then draw from U(0,1.61%]

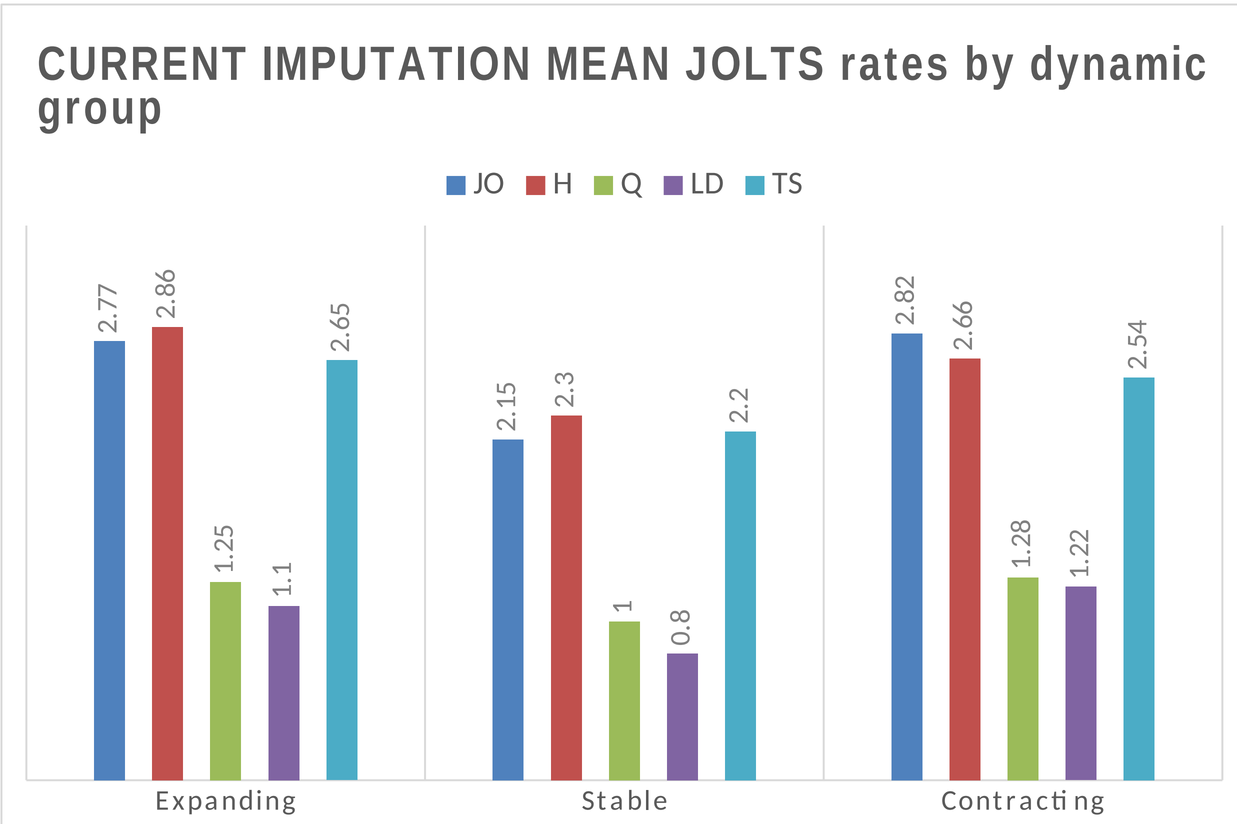
Suppose that the value drawn from U(0,1) is 0.7876, then since 78.76% is larger than β (64%), we will draw from the “ABOVE” distribution, U(1.61%,3.04%). Suppose that the value drawn from U(1.16%,3.04%) is 1.705%. The imputed **rate** for Quits for schedule 70790621 is thus 1.705%, while the imputed level is the reported employment of the recipient establishment multiplied by the imputed rate, so the imputed Quits **level** is 111*1.705%=1.89.

RESULTS

The imputation algorithm described in this paper has been successfully programmed and a complete set of imputed values has been generated on the historical JOLTS series. We can now compare the mean rates of JOLTS respondents, the mean rates generated by the current nearest neighbor algorithm, and the mean rates generated by the proposed alternative. This comparison will be made with respect to employment dynamics.

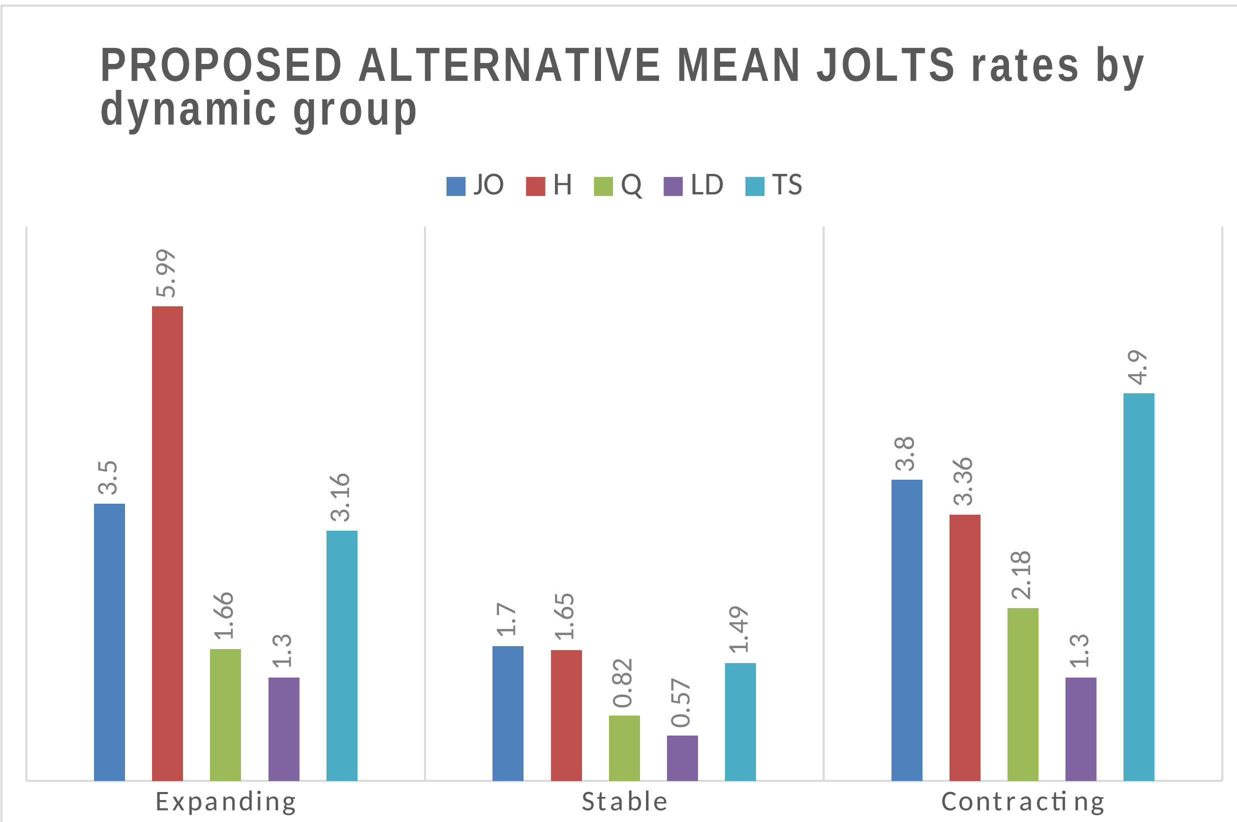


The reported JOLTS data confirms the theoretical relationship detailed previously. Stable establishments tend to have relatively low levels of job openings, hires and separations and, in aggregate, stable establishments tend to have reported hires rates approximately equal to the reported separations rates. Expanding establishments tend to have as a group reported hires rates significantly higher than reported separations rates, and contracting establishments as a group tend to report higher separations rates than hires rates.



The chart above details the mean JOLTS rates for the imputed values generated with the current nearest neighbor algorithm.

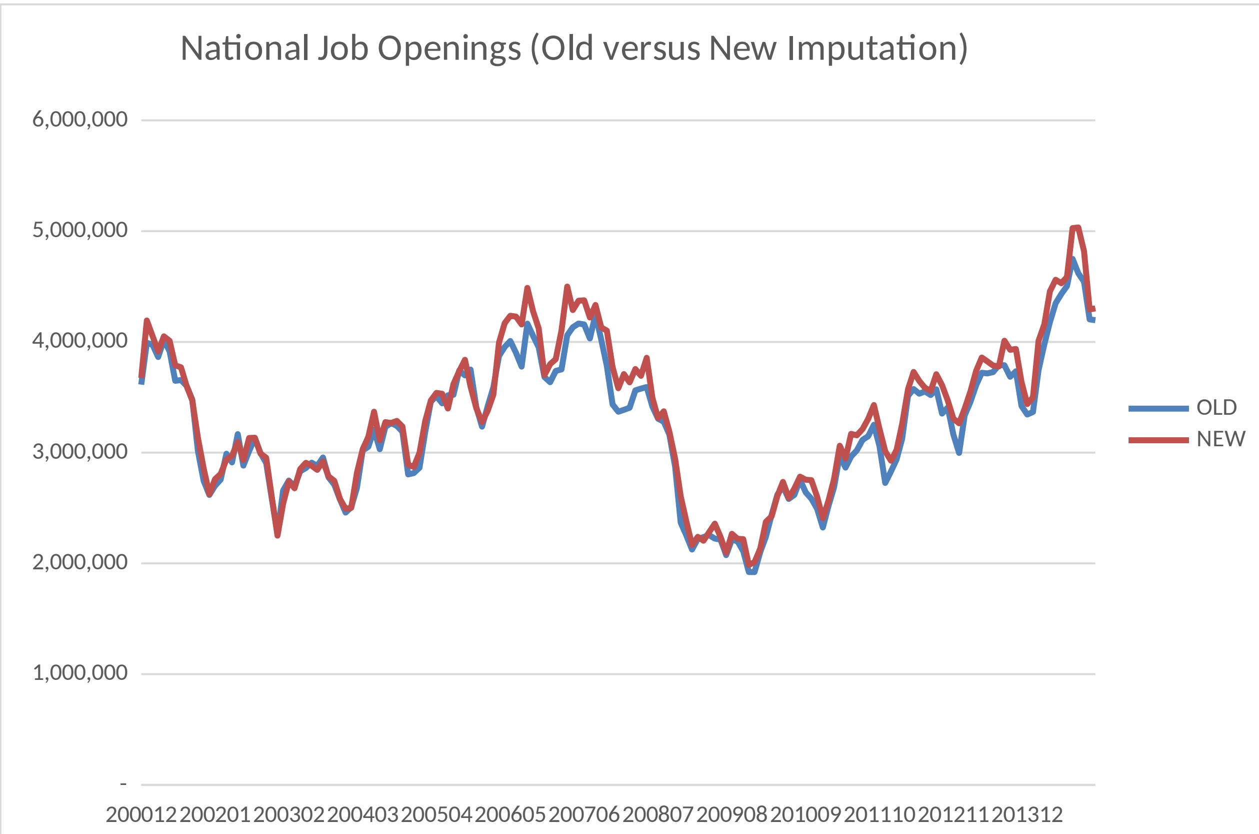
The nearest neighbor imputed JOLTS data is not consistent with the theoretical relationship detailed previously. All of the dynamic groups show the same pattern. That is, expanding establishments, stable establishments, and contracting establishments all have imputed hires rates that are only marginally higher than the imputed separations rates. This inability to match the reported rates of respondents to item non-respondents by dynamic group may contribute to the disequilibrium between the employment growth trend of the Current Employment Statistics (CES) series and the “implied growth” trend of JOLTS (measured as the difference between Hires and Total Separations).



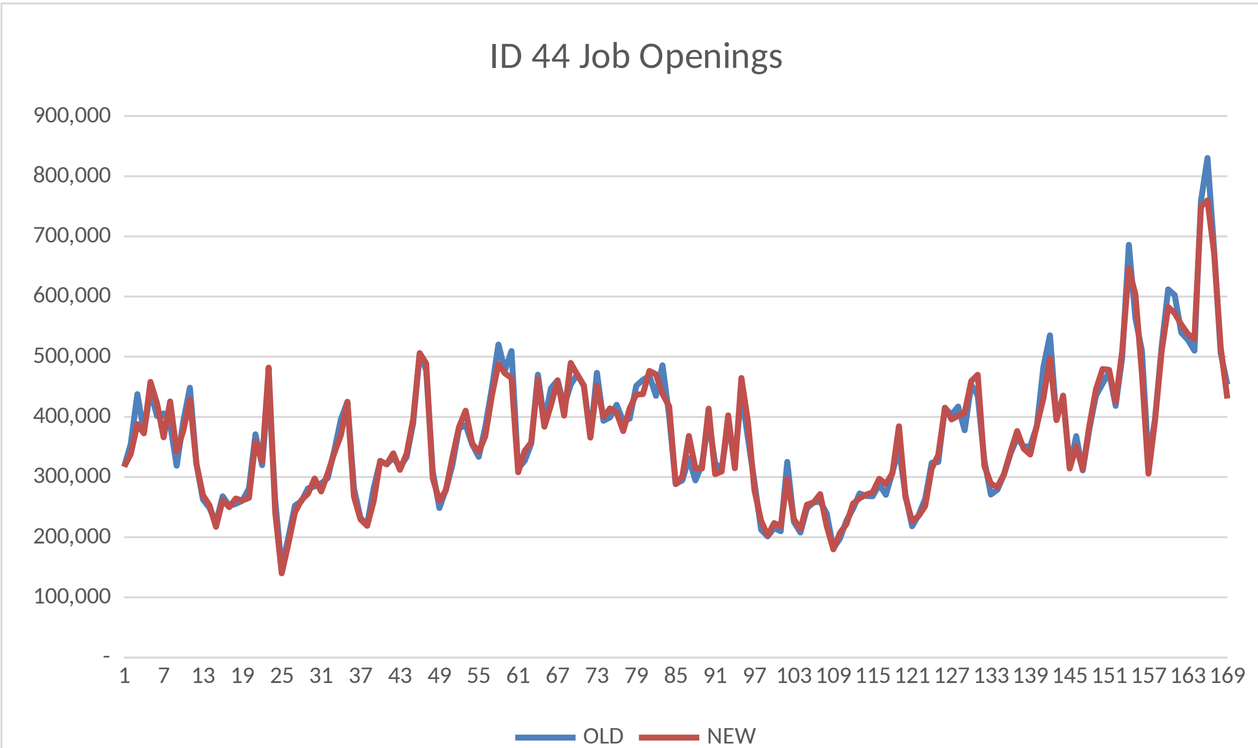
The final chart above details the mean JOLTS rates for the imputed values generated with the proposed alternative. The proposed alternative achieves much better results than the current nearest neighbor algorithm.

The proposed alternative is consistent with the theoretical relationships, although the mean rates of the imputation values generated by the model are generally higher than the actual reported values of item respondents for some variables. (For example, model hires rates in the expanding group are 1% higher than the reported.) Interestingly, the Job Openings rate for contracting establishments is the highest. This could be an indication that firms that have experienced quits or layoffs may be in need of workers and therefore report high levels of job openings. In the proposed model, the distribution lengths (ψ) are derived using an arbitrarily chosen 90% confidence interval of the measure of rate dispersion within the cell (i.e., $\psi = \eta * \sigma$, where η is a constant of 1.645). The rates generated by the model could empirically be made to equal the historical reported rates by altering the value of η for each dynamic group/variable.

The overall effect on JOLTS estimation is modest, primarily owing to the fact that imputation accounts for only 12% of the microdata. Below is the national estimate of Job Openings holding all methodology constant except imputation.



At the industry level, the differences in the estimates using the two imputation approaches generally take the form illustrated in the graph below. There is very little difference in the industry estimates for most months. However, when a significant upward increase in the estimate occurs (i.e., “spikes” in the data), the spike is often slightly smaller when using the new imputation methodology. The truncation of the model has the intended effect of lessening the frequency and magnitude of estimate “spikes” generated solely by imputation.



Summary

The nearest neighbor algorithm currently employed by JOLTS has a number of theoretical and empirical shortcomings. This approach relies solely on the level of reported employment, which carries little JOLTS informative value, and completely disregards employment dynamics, which carries substantial JOLTS informative value. The nearest neighbor approach, owing to the nature of JOLTS rates, is prone to propagating rates that fall far from the cell mean and thus is prone to generate spurious estimation "spikes."

The proposed alternative presented here is intended to be an improvement to the nearest neighbor algorithm. It is not an "off-the-shelf" nor "textbook" application but rather the result of an examination of actual JOLTS rates and an attempt to "mimic" the distribution using simple to understand and easy to derive summary statistics. This attempt to "mimic" JOLTS data was constrained by the conflicting goal to reduce estimation spikes generated by imputation (resulting in a truncated model) and by the narrow scope of my programming ability in an effort to develop, program and test an alternative in a relatively short period of time. This is a preliminary approach that, no doubt, can be improved upon and perhaps, in time, even superseded by a more formal modelling approach. In the meantime, however, the JOLTS estimation series may be improved by the implementation of this alternative with respect to better aligning the rates of imputation reporters and non-reporters with respect to employment dynamics (thus relying less on the CES-JOLTS alignment procedure in JOLTS estimation) and with respect to lessening the frequency and magnitude of estimation spikes solely attributable to imputation (thus allowing the program office to free up production resources now used to investigate imputation- generated spikes).