Estimation Processes Used in the Occupational Requirements Survey October 2015

Bradley D. Rhein[1], Chester H. Ponikowski[1]

[1]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160, Washington, DC 20212

Abstract

The Occupational Requirements Survey (ORS) is an establishment survey conducted by the Bureau of Labor Statistics (BLS) for the Social Security Administration (SSA). The survey collects information on the vocational preparation and the cognitive and physical requirements of occupations in the U.S. economy, as well as the environmental conditions in which those occupations are performed. The survey estimation process consists of multiple steps including calculation of sample weights that reflect a 2-stage stratified sampling design, adjustments for non-response at the establishment level and occupational level, adjustment for collection of data for a unit other than the assigned unit, adjustment for sample rotation, and benchmarking (post-stratification) of weights. Formulas for generating various types of ORS estimates and corresponding estimates of variances also appear in the paper.

Key Words: estimation, establishment survey, sample weights, adjustment factors, benchmarking

1. Introduction

In the summer of 2012, the Social Security Administration (SSA) and the Bureau of Labor Statistics (BLS) signed an interagency agreement, which has been updated annually, to begin the process of testing the collection of data on occupations. As a result, the Occupational Requirements Survey [1] (ORS) was established as a test survey in late 2012. The goal of ORS is to collect and publish occupational information that will replace the outdated data currently used by SSA. More information on the background of ORS can be found in the next section. All ORS products will be made public for use by non-profits, employment agencies, state or federal agencies, the disability community, and other stakeholders.

An ORS interviewer attempts to collect close to 70 data elements related to the occupational requirements of a job. The following four groups of information will be collected:

- Physical demand characteristics/factors of occupations (e.g., strength, hearing, or stooping)
- Specific vocational preparation requirements, which include educational requirements, experience, licensing and certification and post-employment training
- Mental and cognitive demands of work

- Environmental conditions in which the work is completed

The survey plans to publish all estimates that meet the reliability and confidentiality criteria. Somewhere between three and thirteen estimates will be calculated for each of the 70 ORS data elements. Around 700 total estimates could be calculated for a single occupation or occupational group. Estimate types include the percentage of workers in a given category, mean, percentiles (10%, 25%, 50%, 75%, and 90%), and the mode.

This paper explores the ORS estimation processes. Section 2 provides background information on the Occupational Requirements Survey. Section 3 summarizes the ORS sample design. Section 4 summarizes the ORS data elements and types of estimates that will be calculated for each. Section 5 details the process for calculating final sample weights, including the use of non-response adjustment at the unit and quote level, adjustment for collection of data for a unit other than the assigned unit, adjustment for sample rotation, and benchmarking (post-stratification) of weights. Section 6 explores the calculation of standard errors. The paper ends with a conclusion and description of further research to be completed.

2. Background Information on ORS

In addition to providing Social Security benefits to retirees and survivors, the Social Security Administration (SSA) administers two large disability programs, which provide benefit payments to millions of beneficiaries each year. Determinations for adult disability applicants are based on a five-step process that evaluates the capabilities of workers, the requirements of their past work, and their ability to perform other work in the U.S. economy. In some cases, if an applicant is denied disability benefits, SSA policy requires adjudicators to document the decision by citing examples of jobs the claimant can still perform despite restrictions (such as limited ability to balance, stand, or carry objects) [2].

For over 50 years, the Social Security Administration has turned to the Department of Labor's Dictionary of Occupational Titles (DOT) [3] as its primary source of occupational information to process the disability claims [4]. SSA has incorporated many DOT conventions into their disability regulations. However, the DOT was last updated in its entirety in the late 1970's, although a partial update was completed in 1991. Consequently, the SSA adjudicators who make the disability decisions must continue to refer to an increasingly outdated resource because it remains the most compatible with their statutory mandate and is the best source of data at this time.

When an applicant is denied SSA benefits, SSA must sometimes document the decision by citing examples of jobs that the claimant can still perform, despite their functional limitations. However, since the DOT has not been updated for so long, there are some jobs in the American economy that are not even represented in the DOT, and other jobs, in fact many often-cited jobs, no longer exist in large numbers in the American economy.

SSA has investigated numerous alternative data sources for the DOT, such as adapting the Employment and Training Administration's Occupational Information Network (O*NET) [5], using the BLS Occupational Employment Statistics program (OES) [6], and developing their own survey. SSA was not successful with any of these potential data sources and turned to the National Compensation Survey [9] program at the Bureau of Labor Statistics.

3. ORS Sample Design Summary

The ORS sample design is a 2-stage stratified sample of establishments from the entire nation where units are selected by systematic probability proportional to employment size sampling (PPS) at each stage. Stratification is by industry and ownership, directly, and also implicitly by region. Jobs (quotes) are then sampled from the selected establishments by PPS. The frame will be developed from the BLS Quarterly Census of Employment and Wages (QCEW) database [10] with railroads added. Allocation will be proportional to employment size. ORS samples will follow a three-year rotation, though this rotation cycle may change depending on research concerning how often the requirements of work change. For more details on this design, see "Occupational Requirements Survey Sample Design" by Ferguson, et al [7].

4. ORS Data Elements and Possible Estimates Summary

ORS is designed to capture occupational information on educational requirements, cognitive and physical demands, and exposures to environmental conditions. An extensive description of ORS data elements and how estimates for each element will be calculated can be found in the paper "Estimation Considerations for the Occupational Requirements Survey" [8].

Many of the ORS data elements will have the percentage of workers, mean, percentiles, and modes estimates for each occupational definition. For example, one ORS data element measures the amount of time during a typical day that a worker, such as a nurse, spends stooping. Occupational definitions are derived from the Standard Occupational Classification Manual (SOC) [11]. Physical demands, such as stooping, are captured in hours, and so mean and percentile estimates (10%, 25%, 50%, 75%, and 90%) will be calculated. Also, the hours of time spent stooping will fall within an SSA-established category and so a percentage of workers estimate will be calculated for each category. SSA defines five categories by a range of hours spent performing an activity – never, seldomly, occasionally, frequently, and constantly. Finally, the mode of the categories will be identified, marking the twelfth estimate related to stooping.

Estimator type will depend on the type of data element. For any categorical data element, a percentage of workers that fall into a given category will be calculated using Formula 1. The percentage of nurses that are not required to reach overhead or the percentage of security guards that are required to have a bachelor's degree are two examples of categories that will be estimated in ORS. A mode of these percentages will also be identified.

Formula 1 – Calculating Percent of workers

$$\frac{\left[\sum_{i=1}^{I}\sum_{g=1}^{G_i} \text{OccFW}_{ig} \times X_{ig} \times Z_{ig}\right]}{\left[\sum_{i=1}^{I}\sum_{g=1}^{G_i} \text{OccFW}_{ig} \times X_{ig}\right]} \times 100$$

where:

i	= Establishment
I	= Total number of establishments
g	= Occupational quote within establishment i
G_i	= Total number of occupational quotes in establishment i
$OccFW_{ig}$	= Final occupational quote weight (a weight denoting the number of employees in the frame that are represented by occupation g in establishment i) for occupation g in establishment i
X_{ig}	= 1 if occupational quote ig is in the occupational definition
	= 0 otherwise
Z_{ig}	= 1 if the data element is present for occupational quote ig
	= 0 otherwise

For continuous data elements, mean (Formula 2) and percentile (Formula 3) estimates will be calculated. The number of hours that a teacher is required to stand in a typical day or the number of days of previous experience required for a dental hygienist are examples of continuous data elements estimated in ORS. Some continuous data are placed within pre-specified ranges, they would become categorical values, and percentages will be calculated.

Formula 2 – Calculating the Mean

$$\frac{\left[\sum_{i=1}^{I}\sum_{g=1}^{G_i} OccFW_{ig} \times X_{ig} \times Z_{ig} \times Q_{ig}\right]}{\left[\sum_{i=1}^{I}\sum_{g=1}^{G_i} OccFW_{ig} \times X_{ig} \times Z_{ig}\right]}$$

where:
i	= Establishment
I	= Total number of establishments in the survey
g	= Occupational quote within establishment i
G_i	= Total number of occupational quotes in establishment i
$OccFW_{ig}$	= Final occupational quote weight for occupation g in establishment i
X_{ig}	= 1 if occupational quote ig is in the occupational definition
	= 0 otherwise
Z_{ig}	= 1 if the data element is present for occupational quote ig
	= 0 otherwise
Q_{ig}	= Value of the continuous data element for occupational quote ig

Formula 3 – Calculating Percentiles (10%, 25%, 50%, 75%, 90%)

The *p*-th percentile is the value Q_{ig} such that
- the sum of final quote weights ($OccFW_{ig}$) across quotes with a value less than Q_{igj} is less than *p* percent of all final quote weights, and
- the sum of final quote weights ($OccFW_{ig}$) across quotes with a value more than Q_{igj} is less than $(100 – p)$ percent of all final quote weights.

It is possible that there are no specific quotes *ig* for which both of these properties hold. This occurs when there exists a plan for which the $OccFW_{ig}$ of records whose value is less than Q_{ig} equals *p* percent of all final quote weights. In this situation, the *p*-th percentile is the average of Q_{ig} and the value on the record with the next lowest value (note: the SAS proc means procedure handles this special case automatically). The Q_{ig} values must be sorted in ascending order.

Include only quotes that meet the occupational definition and the ORS data element criteria – i.e., where:

$$X_{ig} \times Z_{ig} = 1.$$

where:
i	= Establishment
g	= Occupation within establishment i
X_{ig}	= 1 if quote ig meets the condition set in the occupational definition
	= 0 otherwise
Z_{ig}	= 1 if quote ig meets the condition set in the ORS data element criteria
	= 0 otherwise

5. Derivation of Sample Weights

The calculation of sample weights involves several steps. First, the initial weights are calculated, reflecting the two stages of sampling. Then, the initial weights are adjusted for non-response at the establishment level and occupational level. Next, factors are calculated for the collection of a unit other than the assigned unit, sample rotation, and benchmarking (post-stratification). These factors are applied to the weight that was adjusted for non-response.

Both unit and item non-response are expected to occur in the collection of the Occupational Requirements Survey. Unit non-response occurs when a sampled establishment or occupation refuses to provide any ORS data, while item non-response occurs when we obtain incomplete information for an occupation within an establishment.

ORS plans to adjust for unit non-response by re-weighting establishments and occupations within a pre-specified non-response adjustment cell. Once a weighting adjustment has been made to the establishments and occupations within a cell where a unit refused, the weights of all viable units will be further adjusted by benchmarking the weights to the current national employment to obtain the Final Occupational Quote Weight. Imputation will be used to account for item non-response, though details on this topic are being left for another paper.

Establishment Non-response Adjustment Factor
Non-response adjustment among the establishments involves adjusting the weights of the usable units, those with viable data, to account for the unusable units. The process begins by grouping the establishments into non-response cells, identifying the usable and unusable units, calculating the non-response factor, and then either accepting the factor or collapsing the cell in cases where the factor exceeds a maximum value.

Establishment non-response cells are defined by industry, size class, and area. There are 23 detailed industries, four size classes, and twenty-four areas (see Appendix A). Cells may collapse within industry, size class, and area – that is, a single industry may collapse to a group of industries – before pieces of the cell itself are collapsed. If the cell collapses, area drops out first, then size class. Should the industry collapse pattern be exhausted, any units not adjusted for will remain unadjusted. Collapse patterns can be found in Appendix B.

Non-response cells must collapse under the following circumstances:
- There are zero usable units in the non-response cell
- The non-response adjustment factor exceeds a maximum value

The formula for the non-response adjustment factor follows.

Formula 4 – Non-response Adjustment Factor

$$NRAF = \frac{\sum A + \sum B}{\sum A}$$

where,
$NRAF =$ Non-response Adjustment Factor
$A =$ Weighted Employment for Viable units in the cell
$B =$ Weighted Employment for Non-usable units in the cell

Occupational Non-response Adjustment Factor
Non-response adjustment among occupations follows the same methods as establishment non-response adjustment with one difference – the definition of the non-response cell. For occupations, the non-response cell also includes an occupational group definition (see Appendix C for group definitions and collapse pattern). The occupational group is the last piece of the non-response cell to collapse.

Other Adjustment Factors
Another adjustment factor that may be calculated is the Different than Assigned Factor (DTA factor). In some cases, the collected unit's employment differs greatly from the originally sampled unit. There are two such instances – the collected unit is something more than the originally sampled unit or the collected unit is only a piece of the originally sampled unit. These circumstances are noted by the field and, as a result, a DTA factor will be calculated and applied (the DTA factor is set to 1 otherwise). The formula is as follows.

Formula 5 – DTA Factor

$$DTA\ Factor = \frac{sampled\ employment}{reported\ employment}$$

ORS has a three-year rotation, and each year a fully representative sample will be collected. Since up to three years of data will be used in estimation, a sample group factor must be applied so that each sample group contributes an appropriate amount of the total weighted employment. Given that three equally sized samples are used in estimation, the Sample Rotation Factor for each sample will be 1/3.

There are times when a weight adjustment to an establishment or occupation does not fit into the standard non-response adjustment. Instead, a manual weight adjustment is needed. For example, there may be an ambiguous increase in employment at an establishment where a business fluctuation was noted at the same time as a merger with another company. A manual establishment (or occupational) adjustment factor will allow for such an adjustment to the weight, if necessary. If there are no adjustments, this factor is set to 1.

Benchmarking (Post-stratification) Factor
Benchmarking is the process of adjusting the weights so that the final weighted employment matches a targeted population control number. By matching to a population control number, the estimates of weighted employment will better match the employment levels in the current economy.

The process begins with defining benchmark cells, then a population control number is calculated from the product of a growth factor and the current employment levels before the benchmarking factor is calculated. The original benchmark cells are formed from sampled North American Industry Classification System (NAICS) [12] industry definitions (see Appendix A).

The population control numbers are derived from the QCEW Longitudinal Database (LDB), Railroad file, and the Current Employment Survey (CES). The LDB and Railroad file provide the employment data. Since these sources do not have current employment data, the CES is used to make an adjustment to the employment.

To calculate the employment population control numbers, first a CES Growth Factor is calculated. The CES employment is summed within each benchmark cell for the current quarter and for the reference period of the LDB data. After both sets of CES data are summed by benchmark cell, the following CES Growth Factor for each benchmark cell is calculated.

Formula 6 – CES Growth Factor

CES Growth Factor = Total Current Quarter CES Employment / Total CES Employment for the older reference period

After the CES Growth Factor has been calculated for each cell, the employment from all units in the LDB and OES Railroad file is summed for each benchmark cell. The final employment population control numbers for each cell are calculated by multiplying the total LDB (with railroad file) employment by the CES Growth Factor for the cell.

Formula 7 – Employment Control Number

*Employment control number = LDB employment * CES Growth Factor*

Application of the CES Growth Factor differs slightly for state and local government cells. The CES does not provide data for government at the level of detail of the benchmark cells. The government CES Growth Factors are calculated as described above but are applied to more broadly defined industries. See Appendix A for the details.

During the benchmark process, it may be necessary to collapse benchmark cells and combine the observations into a larger cell. Cell collapsing is the process of combining the

contents of multiple benchmark cells to form a new and larger cell. It is necessary to collapse cells when the number of establishments does not reach a minimum threshold or if the benchmarking factor reaches a maximum value. Collapse patterns define how the cells should be collapsed. See Appendix B for collapse patterns.

The original benchmark cells are collapsed as necessary to achieve the minimum number of establishments per cell. Then, for each collapsed benchmark cell, calculate the total weighted employment.

Formula 8 – Total Weighted Employment by Cell

$$WgtedEmp = \sum_{USE} OccFWgts + \sum_{NMJ} \left(AsgEmp * AsgEstabWgt * SGF \right)$$

where,
$WgtedEmp$ =   Total Weighted Employment
$FinalOccWgt$ =   Final Occupational Quote Weight
$AsgEmp$ =   Sampled Establishment Employment
$AsgEstabWgt$ =   Original Sample Weight
SGF =   Sample Rotation Factor (from Other Factors above)
USE =   All usable quotes
NMJ =   All quotes that out of scope of the survey (No Matching Job)

Next, the employment benchmark factors are calculated for each cell by dividing the employment control number for the cell, calculated in Formula 7, by the total weighted employment derived using Formula 8.

Formula 9 – Employment Benchmark Factor

$$Emp\ Benchmark\ Factor = \frac{Employment\ Control\ Number\ for\ Cell}{Total\ Weighted\ Employment\ for\ Cell}$$

If the employment benchmark factor is less than or equal to the maximum value, then the employment benchmark factor calculation is complete. If the population control number for a cell is 0, then the benchmark factor for the cell is set equal to 1 (A cell with a population control number equal to 0 is not collapsed with other cells.) If the factor is greater than the maximum value, the benchmark cells will collapse to the next level in the collapse pattern and the benchmark factor will be re-calculated for the new cell.

In circumstances for which the benchmark factor is still greater than the maximum value, the process of collapsing cells and calculating benchmark factors continues until all benchmark factors are less than or equal to the maximum value or the collapse patterns are exhausted.

If the collapse patterns are exhausted, and the factor is still greater than the maximum value or if the minimum number of establishments was never met, the value of the benchmark factor will be set to the maximum value for all establishments and occupations in the final cell.

The final quote weight - the weight that represents each sampled occupation in estimation - is the product of the sampling weight and all adjustment factors, shown below in Formula 10.

Formula 10

*Final Occupational Quote Weight ($OccFW_{ig}$) = Sample Weight * Establishment Non-response Adjustment Factor * Occupational Non-response Adjustment Factor * Sample Rotation Factor * DTA Factor * Manual Adjustment Factors * Employment Benchmark Factor*

6. Calculation of Variance Estimates

ORS will provide standard errors for all percentages, means, and percentiles using a modified version of the Balanced Repeated Replication (BRR) method, known as the Fay's BRR method. Using Fay's BRR for variance calculation is a process of several steps. First, variance strata are defined and half-samples are constructed within each of the variance strata definitions. The method requires that the number of replicates be a multiple of four and be at least equal to or greater than the number of variance strata. The ORS sample design specifies 23 detailed industry strata in the private sector and 10 industry strata in the government sector, plus 4 regions, making for 132 variance strata. So, the number of replicates to be used in the calculation of variances will be 132.

A Hadamard Matrix helps construct the replicate half-samples by increasing or decreasing the occupational replicate weight based on the variance strata selection flag. All usable quotes were assigned a selection flag during sampling and will appear in each replicate half-sample with an appropriate weight adjustment.

Once the replicate half-samples are established and the occupational replicate weights have been adjusted, estimation is run for each of the 132 replicate half-samples. Variances are calculated for each estimate, using the sum of the differences between the full sample estimate and each of the 132 replicate estimates. The full sample estimates were calculated using the original, unadjusted weights. The variance formula is as follows:

Formula 11 – Variance Calculation

$$V\left(\hat{\bar{Y}}_c\right) = \frac{1}{R \times (1-k)^2} \sum_{r=1}^{R} \left[\hat{\bar{Y}}_{cr} - \hat{\bar{Y}}_c \right]^2$$

where:

$\hat{V}\left(\hat{\bar{Y}}_c\right)$ = Variance estimate for estimate of ORS data element c

c   = ORS data element

R   = Number of replicates (132)

k   = a constant where $0 \le k < 1$ (the value of k is 0.5)

$\hat{\bar{Y}}_{cr}$ = Estimate for replicate r for ORS data element c

$$\hat{\bar{Y}}_c \qquad\qquad = \text{Estimate for the full sample for ORS data element c}$$

Standard errors are simply the square root of the variance. Relative standard errors are also calculated using the following formula:

Formula 12 – Relative Standard Error Calculation

$$\frac{\sqrt{\hat{V}\left(\hat{\bar{Y}}_c\right)}}{\hat{\bar{Y}}_c} \times 100$$

where,
c $\qquad$ = ORS data element
$\hat{V}\left(\hat{\bar{Y}}_c\right)$ = Variance estimate for estimate of ORS data element c
$\hat{\bar{Y}}_c$ $\qquad$ = Estimate for the full sample for ORS data element c

A significance test will be run to measure the strength of each mode, determining whether the stated mode is significantly different from all of the other possible mode values. A flag will indicate whether the mode is significantly different and, therefore, considered a "strong" mode.

6. Conclusion and Next Steps

ORS will use a few re-weighting methods to account for unit non-response. Weight adjustments will occur at both the establishment and occupational level. Four types of factors will be calculated: non-response adjustment factors, employment different than assigned adjustment factors, manual adjustment factors, and benchmarking (post-stratification) factors.

Non-response adjustment factors will be calculated for cells that contain non-responding units. The weight of responding units will increase to account for the weight of non-responding units. Collapsing of the cell will occur if no responding units can be found in the cell, or if the non-response adjustment factor exceeds a pre-determined maximum value. These factors will be calculated for both establishments and occupations.

A different than assigned (DTA) factor will be calculated for all establishments. This factor adjusts for differences in the assigned sample unit compared with the collected sample unit. There are two ways this could happen: an establishment could be collected with employment larger than assigned or it could be collected with employment less than assigned. In either case, a weight adjustment will be made to account for the difference in employment.

Manual adjustment factors are calculated when an adjustment needs to be made but such an adjustment does not fit within the non-response adjustment factor or DTA factor. In these circumstances, a special factor is calculated manually. Such a factor could be calculated for establishments or occupations.

Benchmarking factors ensure that the final weighted employment matches the current employment in the economy. For this process, a ratio will be calculated comparing the frame employment at the time of sampling to employment currently found in the economy. Benchmarking factors are applied to all weights that are used in the estimation process.

ORS will provide a measure of accuracy for every estimate produced. For means, percentages, and percentiles, a standard error will be calculated using Fay's BRR method. For modes, a significance test will be performed to determine if the mode is significantly different from the other possible values.

ORS plans to adjust for item non-response by using imputation methods or providing estimates for missing or unknown responses. Methods are currently being investigated. Testing began in the spring of 2015 and is ongoing. Results will be presented in a future paper.

References/Footnotes

[1]   Occupational Requirements Survey, http://www.bls.gov/ors/.
[2]   Social Security Administration, Occupational Information System Project, http://www.ssa.gov/disabilityresearch/occupational_info_systems.html.
[3]   U.S. Department of Labor, Employment and Training Administration (1991), "Dictionary of Occupational Titles, Fourth Edition, Revised 1991"
[4]   Occupational Information Development Advisory Panel, 2010, http://www.socialsecurity.gov/oidap/index.htm
[5]   U.S. Department of Labor, O*Net Online, http://www.onetonline.org/
[6]   Bureau of Labor Statistics, Occupational Employment Statistics Program, http://www.bls.gov/oes/
[7]   Ferguson, Gwyn R., McNulty, Erin. 2015. Occupational Requirements Survey Sample Design. In *JSM proceedings*, Economic Data: CPI, PPI, NCS Section. Alexandria, VA: American Statistical Association.
[8]   Rhein, Brad, Ponikowski, Chester, and McNulty, Erin. 2014. Estimation Considerations for the Occupational Requirements Survey. In *JSM proceedings*, Government Statistics Section. Alexandria, VA: American Statistical Association. 2134-2146.
[9]   National Compensation Survey, http://www.bls.gov/ncs/.
[10]   Quarterly Census of Employment and Wages, http://www.bls.gov/cew/.
[11]   Standard Occupational Classification System, http://www.bls.gov/soc/.
[12]   North American Industry Classification System, http://www.census.gov/eos/www/naics/.

Appendix A – Industry, Size Class, and Area Definitions

Private Industry

Cell	NAICS Codes	Industry
21A	21	Mining
23A	23	Construction
31A	31-33 (including 336411)	Manufacturing
22A	22	Utilities
42A	42	Wholesale Trade
44A	44-45	Retail Trade
48A	48-49	Transportation and Warehousing
51A	51	Information
52A	52(excl 524)	Finance (Rest of)
52B	524	Insurance
53A	53	Real Estate, Renting, Leasing
54A	54	Professional and Scientific Services
55A	55	Management of Companies/Enterprises
56A	56	Admin., Support, Waste Management
61A	61(excl 6111-6113)	Educational Services (Rest of)
61B	6111	Elementary and Secondary Schools
61C	6112,6113	Junior Colleges, Colleges & Universities
62A	62(excl 622,623)	Health Care, Social Assistance (Rest of)
62B	622	Hospitals
62C	623	Nursing and Residential Care Facilities
71A	71	Arts, Entertainment, Recreation
72A	72	Accommodation and Food Services
81A	81(excl 814)	Other Services (except Public Admin.)

Local Government

Cell	NAICS Codes	Industry
10L	21, 23, 31-33	Goods-Producing
20L	42, 44-45, 48-49, 22	Trade, Transportation, and Utilities
30L	6111	Elementary and Secondary Education
40L	6112, 6113	Colleges and Universities
50L	61 excl. 6111-6113	Rest of Education
60L	622	Hospitals
70L	623	Nursing Homes
80L	62 excl. 622-623	Rest of Health and Social Services
90L	92 excl. 928	Public Administration
99L	51, 52-53, 54-56, 71-72, 81	Other Service-producing

State Government

Cell	NAICS Codes	Industry
10S	21, 23, 31-33	Goods-Producing
20S	42, 44-45, 48-49, 22	Trade, Transportation, and Utilities
30S	6111	Elementary and Secondary Education
40S	6112, 6113	Colleges and Universities
50S	61 excl. 6111-6113	Rest of Education
60S	622	Hospitals
70S	623	Nursing Homes
80S	62 excl. 622-623	Rest of Health and Social Services

90S	92 excl. 928	Public Administration
99S	51, 52-53, 54-56, 71-72, 81	Other Service-producing

Size Class	Employment Range
1	0-49
2	50-99
3	100-499
4	500+

Area	Area Code
Atlanta-Athens-Clarke-Sandy Springs, GA CSA	00122
Boston-Worcester-Providence, MA-NH-RI-CT CSA	00148
Chicago-Naperville, IL-IN CSA	00176
Dallas-Fort Worth, TX CSA	00206
Detroit-Warren-Ann Arbor, MI CSA	00220
Houston-The Woodlands, TX CSA	00288
Los Angeles-Long Beach, CA CSA	00348
Minneapolis-St. Paul, MN-WI CSA	00378
New York-Newark, NY-NJ-CT-PA CSA	00408
Philadelphia-Camden-Reading, PA-NJ-DE-MD CSA	00428
San Jose-San Francisco-Oakland, CA CSA	00488
Seattle-Tacoma, WA CSA	00500
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA	00548
Miami-Fort Lauderdale-Port St. Lucie, FL	33100
Phoenix-Mesa-Scottsdale, AZ	38060
Rest of New England	99901
Rest of Middle Atlantic	99902
Rest of East South Central	99903
Rest of South Atlantic	99904
Rest of East North Central	99905
Rest of West North Central	99906
Rest of West South Central	99907
Rest of Mountain	99908
Rest of Pacific	99909

Appendix B – Collapse Patterns for Industry, Size Class, and Area

Industry	NAICS Codes	Collapse Pattern
21A	21	none
22A	22	22A&48A
23A	23	none
31A	31-33 (including 336411)	none
42A	42	none
44A	44-45	none
48A	48-49	22A&48A
51A	51	51A-53A
52A	52(excl 524)	51A-53A
52B	524	51A-53A

53A	53	51A-53A
54A	54	54A-56A
55A	55	54A-56A
56A	56	54A-56A
61A	61(excl 6111-6113)	none
61B	6111	none
61C	6112,6113	none
62A	62(excl 622,623)	none
62B	622	none
62C	623	none
71A	71	71A-81A
72A	72	71A-81A
81A	81(excl 814)	71A-81A
10L	21, 23, 31-33	10L,20L,90L,99L; 10L-99L
20L	42, 44-45, 48-49, 22	10L,20L,90L,99L; 10L-99L
30L	6111	30L,40L,50L; 10L-99L
40L	6112, 6113	30L,40L,50L; 10L-99L
50L	61 excl. 6111-6113	30L,40L,50L; 10L-99L
60L	622	60L,70L,80L; 10L-99L
70L	623	60L,70L,80L; 10L-99L
80L	62 excl. 622-623	60L,70L,80L; 10L-99L
90L	92 excl. 928	10L,20L,90L,99L; 10L-99L
99L	51, 52-53, 54-56, 71-72, 81	10L,20L,90L,99L; 10L-99L
10S	21, 23, 31-33	10S,20S,90S,99S; 10S-99S
20S	42, 44-45, 48-49, 22	10S,20S,90S,99S; 10S-99S
30S	6111	30S,40S,50S; 10S-99S
40S	6112, 6113	30S,40S,50S; 10S-99S
50S	61 excl. 6111-6113	30S,40S,50S; 10S-99S
60S	622	60S,70S,80S; 10S-99S
70S	623	60S,70S,80S; 10S-99S
80S	62 excl. 622-623	60S,70S,80S; 10S-99S
90S	92 excl. 928	10S,20S,90S,99S; 10S-99S
99S	51, 52-53, 54-56, 71-72, 81	10S,20S,90S,99S; 10S-99S

Size Class	Employment Range	Collapse Pattern
1	0-49	1; 1-2; 1-4
2	50-99	2; 1-2; 1-4
3	100-499	3; 3-4; 1-4
4	500+	4; 3-4; 1-4

Area #	Area Name	Area Cell	1st Collapse	2nd Collapse
1	Atlanta	00122	1, 13, 14	1, 13, 14, 19
2	Boston	00148	2, 16	none
3	Chicago	00176	3, 5	3, 5, 20
4	Dallas	00206	4, 6	4, 6, 22
5	Detroit	00220	3, 5	3, 5, 20
6	Houston	00288	4, 6	4, 6, 22
7	Los Angeles	00348	7, 11, 12	7, 11, 12, 24
8	Minneapolis	00378	8, 21	none

9	New York	00408	9, 10	9, 10, 17
10	Philadelphia	00428	9, 10	9, 10, 17
11	San Francisco	00488	7, 11, 12	7, 11, 12, 24
12	Seattle	00500	7, 11, 12	7, 11, 12, 24
13	Washington	00548	1, 13, 14	1, 13, 14, 19
14	Miami	33100	1, 13, 14	1, 13, 14, 19
15	Phoenix	38060	15, 23	none
16	Rest of New England	99901	2, 16	none
17	Rest of Middle Atlantic	99902	9, 10, 17	none
18	East South Central	99903	none	none
19	Rest of South Atlantic	99904	1, 13, 14, 19	none
20	Rest of East North Central	99905	3, 5, 20	none
21	Rest of West North Central	99906	8, 21	none
22	Rest of West South Central	99907	4, 6, 22	none
23	Rest of Mountain	99908	15, 23	none
24	Rest of Pacific	99909	7, 11, 12, 24	none

Appendix C – SOC Aggregate Group (SAG) Definitions

SAG	SOC Code	SOC
1	11, 13	Management, Business, Financial
2	15, 17, 19, 21, 23, 25, 27, 29	Professional and Related
3	31, 33, 35, 37, 39	Service
4	41	Sales and Related
5	43	Office and Administrative Support
6	45, 47	Construction, Extraction, Farming, Fishing
7	49	Installation, Maintenance, and Repair
8	51	Production
9	53	Transportation and Material Moving