Evaluation of Alternative Measures of Size for Sampling of Establishments in the NCS October 2012

Bradley D. Rhein[1], Chester H. Ponikowski[2], and Joan Coleman[3]
[1]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160,
Washington, DC 20212
[2]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160,
Washington, DC 20212
[3]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160,
Washington, DC 20212

Abstract
The National Compensation Survey, conducted by the Bureau of Labor Statistics, is an establishment survey sampled yearly from a national frame using probability proportionate to establishment employment size. The national frame is developed from administrative files maintained quarterly by the States for the Unemployment Insurance program. Each establishment on the frame is assigned a measure of employment size equal to the employment in the third month of the frame quarter. In 2011, approximately 15% of the establishments in the frame, which includes seasonal businesses, reported an employment of zero, an employment that must be adjusted to ensure that the establishment has a chance of selection. In the past, establishments with zero employment have been assigned an employment equal to one employee, with some cases resulting in large weights for the occupations selected from these companies. The large weights lead to an over-representation of these occupations in the sample. This paper presents alternative measures of size values and discusses the options for best determining the measure of size for all establishments in the frame.

Key Words: sampling frame, establishment survey, sample selection, weight

1. Introduction

Sampling proportional to measure of size is known to be an efficient method when the measure of size is nearly proportional to the unit totals of principal survey items. An efficient method is one that results in lower sampling variance than other commonly used methods such as simple random sampling or stratified simple random sampling. One of the principal items in the National Compensation Survey, conducted by the Bureau of Labor Statistics, is the employer total compensation cost per employee hour worked. The NCS data show that employer compensation costs increase with the size of the establishment. Therefore, it is desirable for NCS to use sampling proportional to establishment employment. However, to sample with probability proportional to employment size, every establishment included in the sampling frame needs to have a positive employment. But some establishments in the NCS sampling frame have zero employment because they are seasonal units, newly forming, or have ceased operations. These establishments with a zero employment do not have a chance of being in the sample unless their employment is changed to a positive employment. The key question being explored in this paper is what positive employment value to use for establishments in the frame with an employment of zero.

In March of 2011, approximately 15% of the 7,727,581 establishments in the frame for private industry reported an employment of zero. In the past, establishments with zero employment have been assigned an

employment measure of size equal to one employee. In some situations the actual employment turned out to be much greater than one, resulting in large weights for the occupations selected from these establishments. The large weights lead to an over-representation of these occupations in the sample.

This paper reports an analysis of the frame establishments with zero employment, identifies nine alternative measures of size for these units, and evaluates these nine alternative measures using simulated samples and available frame employment and earnings data. Section 2 provides an overview of the NCS program and sample design. Section 3 provides an analysis of frame establishments with zero employment. Section 4 presents the alternative measures that were included in the study. Section 5 describes the simulation study while Section 6 presents the simulation results. Finally, Section 7 presents our conclusion and topics for further research.

2. Overview of the NCS

The NCS provides comprehensive measures of employer costs for employee compensation, compensation trends, and the incidence and provisions of employer-provided benefits.

The NCS produces several types of data with varying degrees of frequency as summarized below.
- Employment Cost Index (ECI) data are released quarterly
- Employer Costs for Employee Compensation (ECEC) data are released quarterly
- Incidence and Provisions of Employer Provided Benefits data are released annually
- Detailed Provisions for employer provided health insurance, defined benefit retirement plans, and defined contribution retirement plans are released once a year with a focus on one of these benefit areas each year

The NCS covers workers in private industry establishments and in State and local government for all 50 States and the District of Columbia. Establishments with one or more workers are included in the survey scope. Excluded from the survey are workers in the Federal Government and quasi-Federal agencies, military personnel, agricultural industry, workers in private households, the self-employed, volunteers, unpaid workers, individuals receiving long-term disability compensation, individuals working overseas, individuals who set their own pay (for example, proprietors, owners, major stockholders, and partners in unincorporated firms), and those paid token wages.

The BLS Quarterly Census of Employment and Wages (QCEW) serves as the sampling frame for the NCS sample. The QCEW is created from State Unemployment Insurance (UI) files of establishments, which are obtained through the cooperation of the individual state agencies (BLS Handbook of Methods, Chapter 5). This sampling frame includes many useful pieces of data for NCS including monthly employment counts for each establishment, total quarterly wages for the establishment, establishment identification data and contact information. The QCEW sampling frame includes all establishments, including units with monthly employment that are consistently positive, some with seasonal employment, newly formed businesses that may not yet have any employees, and establishments that have recently ceased operations. All establishments with one or more employees at any time during the year before the initiation of an NCS sample are considered to be in-scope for the NCS.

Recently, the NCS has undergone a sample redesign. The redesigned NCS sample consists of three rotating replacement sample panels for private industry establishments, an additional sample panel for State and local government entities, and an additional panel for private industry firms in the aircraft manufacturing industry. Each of the sample panels is in the sample for at least three years before it is replaced by a new sample panel from the most current frame. Establishments in each sample panel are

initiated over a 15-month time period. After initiation, data are updated quarterly for each selected establishment and occupation until the panel in which the establishment was selected is replaced. Estimates for all outputs, except Detailed Provisions, use data from the entire set of five independent sample panels.

The redesigned NCS sample is selected using a two stage stratified design with probability proportionate to employment size (PPS) sampling at each stage. The first stage of sample selection is a probability sample of establishments in 24 pre-determined geographic area strata and 5 aggregate industries. The second stage is a probability selection of occupations (PSO) within the establishments. The 24 areas consist of the 15 largest metropolitan areas by employment and the rest of each of the nine Census Divisions, excluding the 15 largest areas. A more detailed description of the new NCS sample design is given in Ferguson, et al. (2011) while a description of the estimates produced and the estimate methodology is given in Chapter 8 of BLS Handbook of Methods.

3. Establishments with Zero Employment in the Sampling Frame

Currently, NCS extracts one calendar quarter of QCEW data for use as the sampling frame, choosing the third month of the quarter to be the reference month. For this analysis, four calendar years of data from 2007 to 2011 were extracted from QCEW database, providing the following data elements: identification, monthly employment, total quarterly wages, NAICS code, location, and ownership. Private industry data were the focus for this study and units with zero employment in December 2007 serve as the base for all analysis. December of 2007, then, serves as the reference month and the calendar year of 2007 serves as the reference year.

In December of 2007 there were 988,264 establishments with zero employment out of 7,843,365 establishments (about 12.6%) on the sampling frame. Such establishments are referred to as zero employment establishments (ZEE). While these units exist on the frame, they will have no chance of selection without applying an alternative measure of establishment size.

There are at least three reasons why an establishment would have zero employment on the frame; an establishment could be 1) a newly forming unit 2) a seasonal unit or 3) a unit that is going out of business.

- A newly forming unit was defined as a unit that did not have positive employment at any point in the reference year. Newly forming units have acquired a business account but were not yet required to submit a payroll report to the Unemployment Insurance Program.

- Seasonal units are zero employment establishments that had at least one month of positive employment in the reference year, yet had zero employment in the reference month. These units may have predictable months of employment and zero employment.

- Units with positive employment in the reference year that dropped to zero prior to or during the reference month could be establishments that have ceased operations – the unit was going out of business and no longer had payroll to report. These units are difficult to separate from the seasonal unit definition.

Of the 988,264 zero employment establishments, 404,844 were newly forming units – about 5.2% of the frame – and 583,420 were seasonal establishments or units that ceased operations – about 7.4% of the frame.

Since frame data is available from 2007 through 2010, the employment and wage trends of 2007 zero employment establishments can be studied. Chart 1 below illustrates the employment of these units in 2008, when data from sampled zero employment establishments from 2007 would be collected. Out of all zero employment establishments from 2007, 8% have consistent positive employment in 2008, while 21% have some months of positive employment and some months of zero employment. The rest, 71%, either continue to have zero employment or do not appear on the frame in 2008.
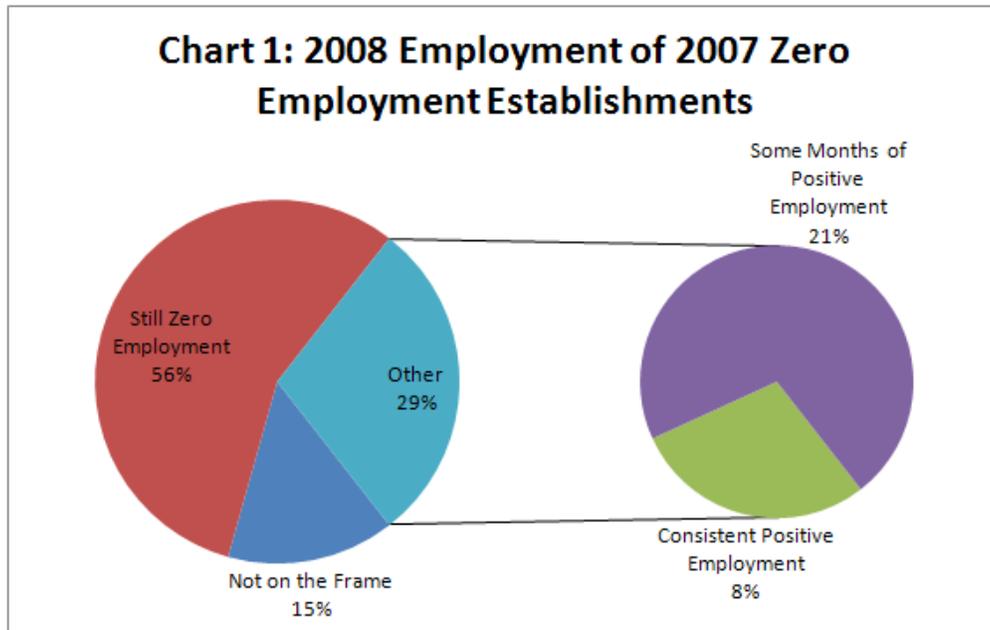
Finding positive employment for a 2007 zero employment establishment is encouraging, but the unit will need to have both positive employment and wage data in 2010 to eventually contribute to estimation. Chart 2 illustrates the status over time for newly forming units and seasonal units from the 2007 frame.
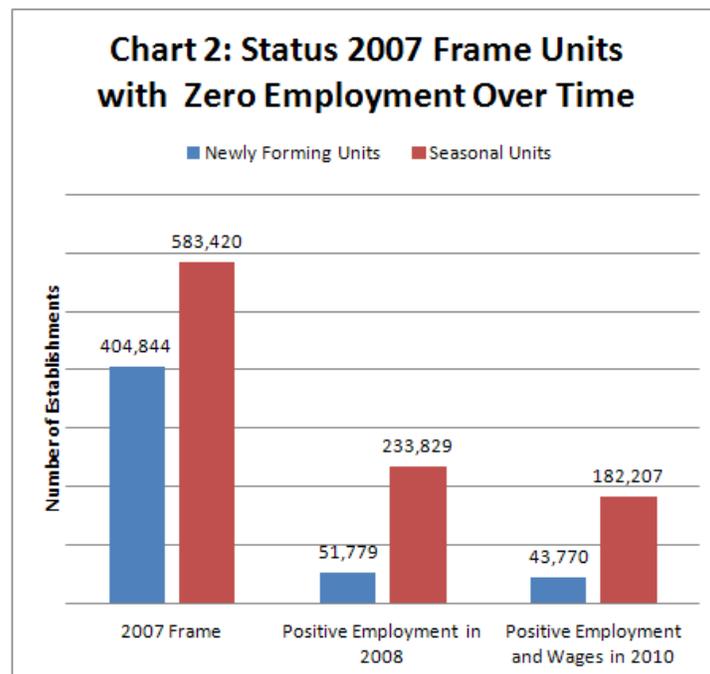
Approximately 12% of the newly forming establishments and 40% of the seasonal units from 2007 have positive employment in 2008. Once a zero employment establishment has positive collected employment, the percentage of units having a positive employment and eligible to contribute to estimation in 2010 is at most 85% for newly forming units and 78% for seasonal units. The eligible units are subject to refusal or non-contact during collection.

However, the graph shows that of the 404,844 newly forming units existing in 2007, only 10% had a positive employment and could contribute to estimation in 2010. Seasonal establishments fare better, having about 31% eligible to contribute to estimation in 2010. These rates of contribution are much less than the rate for establishments that report positive employment in the reference month (80%). Overall, when considering the proportion of frame units that could contribute to estimation in 2010, newly forming and seasonal units make up 0.6% and 2.3% of the frame, respectively. These low rates suggest that newly forming and seasonal units will be less likely to appear in the sample, have positive employment when collected, and therefore be eligible for contributing to estimation.

Newly forming establishments demand more attention because there is no employment information for these units over a period of 12 months. In 2007, the sampling frame consisted of 404,844 newly forming units, 12.8% of which had positive employment in 2008. Chart 3 shows the distribution of 2008 employment for these newly forming establishments by percentage and number of employees.
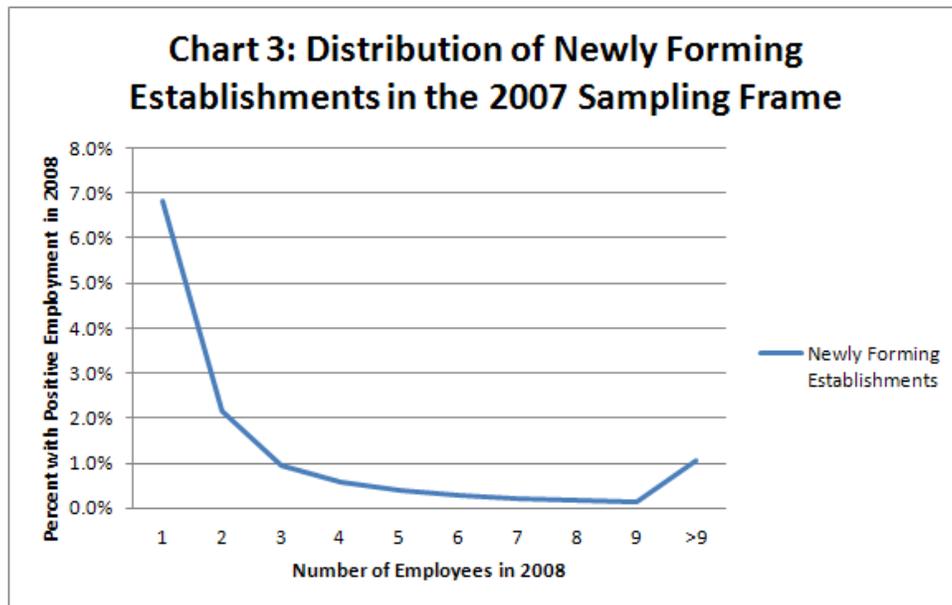
The graph shows that about 9% of the newly forming units have one or two employees in 2008. Given that 12.8% of the newly forming units had positive employment, the percentage of units with one or two employees is nearly 75% for newly forming establishments with positive employment. At the other tail, one percent of the newly forming units have more than 9 employees. These units are likely to have large weights since the measure of size will under-represent the establishment, leading to a large sampling weight. The combination of a large sampling weight and a large employment produces weights that are much larger than desired. Chart 4 breaks out the employment distribution by aggregate industry.

Chart 4: Distribution of Newly Forming Establishments in the 2007 Sampling Frame by Industry
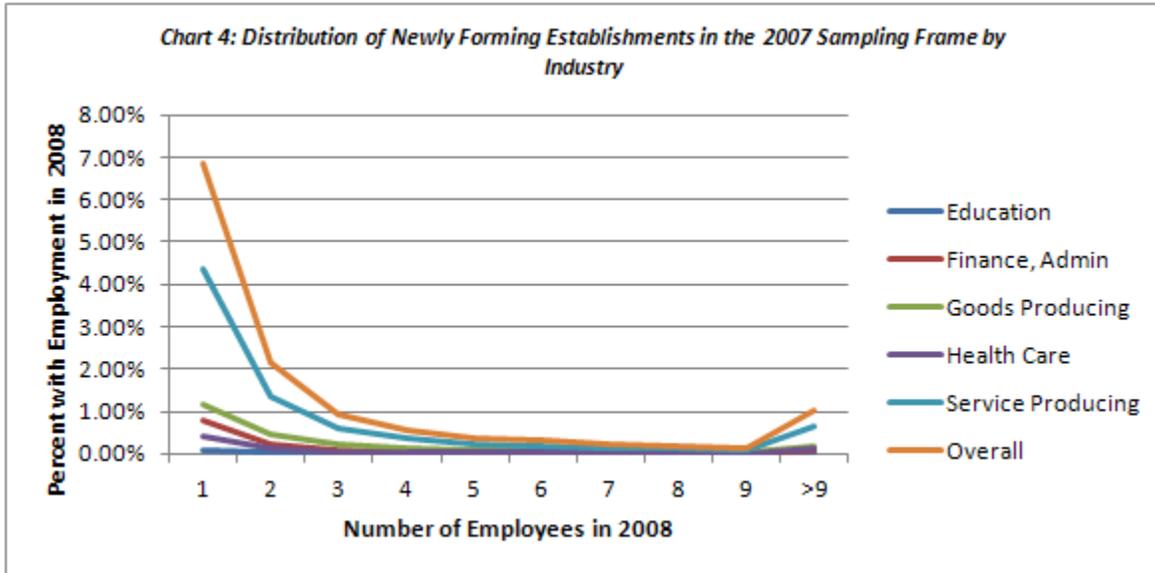
4. Options for Establishment Measure of Size

This research will cover three suggested alternative approaches to setting the measure of establishment size as described in the list below – set employment to a fixed value, an average, or a maximum. All measures of establishment size assignments must be greater than zero for all units on the sampling frame.

Each measure of establishment size approach described here was applied only to units with zero employment in the reference month. For all other establishments in the frame, the employment in the reference month was assigned as the measure of establishment size.

1. Use a Fixed Value for the Measure of Size
    a. MOS_1 – Establishments are assigned one employee
    b. MOS_4 – Establishments are assigned four employees
    c. MOS_8 – Establishments are assigned eight employees
2. Use an Average for the Measure of Size
    a. MOS_AVE – Establishments are assigned an average employment by extracting a full year of data and calculating the average positive employment within those 12 months of data. For those establishments with no positive employment over the 12 month period each of the following values were explored:
        i. MOS_AVE (1) – if the average is zero, then set the employment to 1
        ii. MOS_AVE (4) – if the average is zero, then set the employment to 4
        iii. MOS_AVE (8) – if the average is zero, then set the employment to 8
3. Use a Maximum for the Measure of Size
    a. MOS_MAX – Establishments are assigned an employment equal to the 12-month maximum employment by extracting a full year of data and using the maximum employment over the 12 month period. For those establishments with no positive employment over the 12 month period each of the following values were explored:
        i. MOS_MAX (1) – if the maximum is zero, then set the employment to 1
        ii. MOS_MAX (4) – if the maximum is zero, then set the employment to 4
        iii. MOS_MAX (8) – if the maximum is zero, then set the employment to 8

5. Simulation Study: Analysis of Establishments with Zero Employment in Sampling

The goal of the simulation study was to assess the performance of the different measures of employment size by analyzing the sampling of zero employment units and then comparing the sample estimates of average monthly wages to frame estimates. For the purposes of this research, various data elements were taken from different points in time with an attempt to mirror what happens in the lifecycle of an establishment selected in the NCS. A frame from 2007 was used for sampling, employment from 2008 was used as a proxy for collected employment, and monthly wages from 2010 were used to make a comparison of the future performance of each measure of establishment size used during sampling.

Total quarterly wages, found on the QCEW frame, were converted to monthly wages by dividing the total quarterly wages by the sum of monthly employment for the quarter. Establishments with zero employment in each month of the quarter were given average monthly wages of zero. Estimates of average monthly wages were produced for aggregate industries, detailed industries, and ECI localities as defined in Appendix A.

Two hundred simulated samples were drawn from the third month of the fourth quarter frame of 2007 using the various measures of size defined earlier. Sampling was completed using current NCS sampling methods. NCS is a national survey sampled with probability proportional to establishment size within the 120 area/industry strata, comprising 24 areas and 5 aggregate industries. The sampling also involves an implicit sampling of 23 detailed industry strata within each of the areas. A listing of areas, aggregate industries, and detailed industries appears in Appendix A. Allocation of the sample depends on the sample size - 9,754 establishments - and the dispersion of employment by industry throughout the sampling frame.

Table 1 below shows the results of sampling zero employment establishments (ZEE) for each alternative measure of size. More zero employment units are selected and found to be contributing to estimation with larger fixed values assigned as the establishment measure of size. However, using smaller fixed values results in both higher percentages of zero employment establishments with positive employment in 2008 and higher rates of establishments eligible to contribute to estimation in 2010. Also, since fewer zero employment establishments are selected, fewer units are lost, minimizing wasted allocation. For all alternative measures of size, the average percentage of sampled zero employment units that still reported zero employment in 2008 is within the range of 77% to 90%.

Table 1: Summary of Zero Employment Establishments Sampled over 200 Simulated Samples

Measure of Size	Average Number of ZEE per Sample	Average Percentage of ZEE per Sample	Average Percentage of ZEE per Sample that had Positive Employment in 2008	Average Percentage of ZEE per Sample eligible to contribute to Estimation in 2010	Number of Sampled Units Lost in Collection due to Zero Employment
MOS_1	91	0.9%	17.4%	11.9%	80
MOS_4	357	3.7%	17.2%	12.0%	314
MOS_8	685	7.0%	17.1%	11.7%	605
MOS_AVE(1)	208	2.1%	21.2%	15.9%	175
MOS_AVE(4)	313	3.2%	16.6%	12.1%	275
MOS_AVE(8)	453	4.6%	13.7%	9.7%	409
MOS_MAX(1)	252	2.6%	23.0%	17.1%	209
MOS_MAX(4)	358	3.7%	18.2%	13.4%	310
MOS_MAX(8)	494	5.1%	15.2%	10.9%	440

	Number of ZEE	Percentage of ZEE	Percentage of ZEE with Positive Employment in 2008	Percentage of ZEE that could contribute to Estimation in 2010	Number of ZEE still with Zero Employment in 2010
2007 Frame ZEE	988,264	12.6%	28.9%	22.9%	761,952

When a unit's collected employment was 1 or 2, MOS_1 would be the most accurate measure of size to use. Similarly, for units with an employment of (3, 4, 5) or (6, 7, 8, 9), the ideal measure of size would be MOS_4 and MOS_8, respectively. The following graph illustrates the distribution of reported employment from 2008 for the newly forming units that were sampled during the simulation study.

Chart 5: Distribution of the 2008 Average Collected Employment for Newly Forming Establishments over 200 Samples
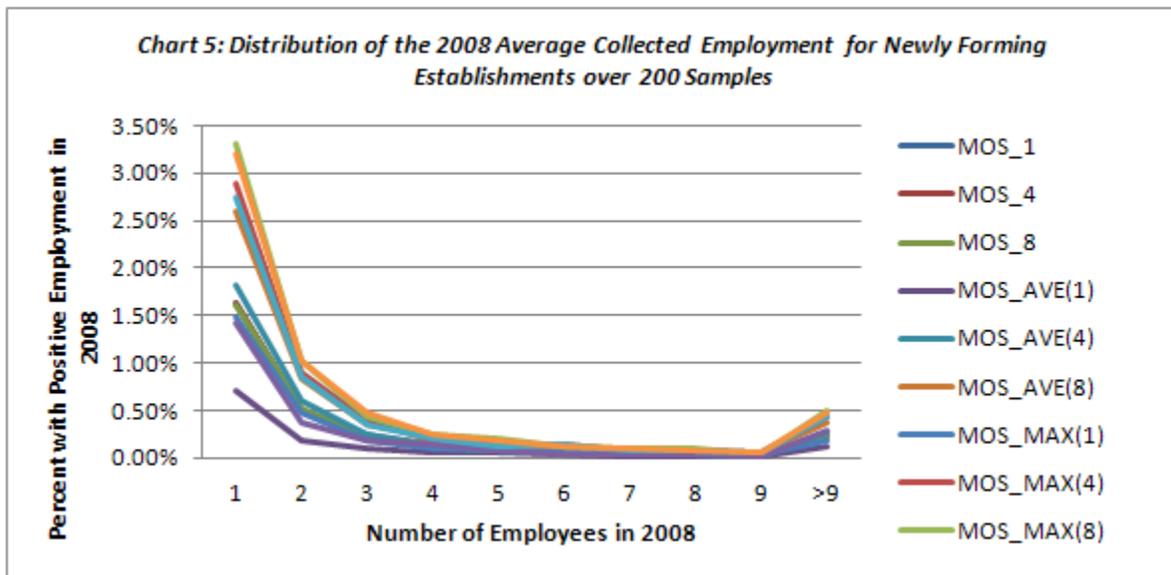
The results show that most sampled newly forming units have a collected employment of 1, 2, 3, or something greater than 9 employees. However, the number of newly forming units actually sampled was

very low, especially when assigning a fixed value of 1. While using MOS_1 resulted in sampling an average of two or three units, assigning MOS_4 and MOS_8 resulting in 10 and 20 sampled newly forming establishments, respectively. For MOS_AVE (8), where newly forming units were set to 8 employees, 22 establishments with positive employment were selected, 17 of which had one or two employees. In this case, the measure of size equal to 8 over-represents 17 out of 22 newly forming establishments in sampling, leading to sampling weights that, at least initially, under-represent the true size of the unit.

Reaching the goal of assigning an appropriate measure of establishment size to the zero employment establishments in the frame is a difficult task. Sampling many of these units tends to result in a higher number of contributing establishments, but also a higher rate of loss. For newly forming units, using a fixed value of 1 tends to result in high rates of contribution to estimation and low loss rates, but few of these establishments will appear in the sample. Regardless of the alternative measure of size used, the average sample loss rate, in terms of units able to contribute to estimation, for zero employment establishments is within the range of 83% to 91%.

6. Results of the Simulation Study for the Units with Zero Employment

Average monthly wages were calculated for three types of estimation cells from 1) the 200 sample simulations and 2) the sampling frame. Once calculated, the average monthly wages from the samples could be compared to the frame results. Wages from the frame served as expected values for the wages calculated from the simulated samples. Along with comparing the full sample to the full frame, a subset consisting of only the zero employment establishments from each sample was compared to a frame of zero employment establishments.

First, the value was computed for area-industry cells that combine the 5 aggregate industries with the 24 areas, including the 15 locality ECI areas. Second, the average monthly wage was calculated for 23 detailed industry cells, including industries such as retail trade, elementary schools, and hospitals. The last type of estimation cell was aggregate industry. The areas, aggregate industries, and detailed industries are listed in Appendix A.

Once all average monthly wages were calculated, a comparison could be made between the samples and the frame. Aggregate industry cells were the only cell definition that had stable results; the other cell definitions were too specific, having either too many extreme average monthly wages or numerous cells that did not have any data for comparison.

First, the average monthly wage by cell was established for the frame of zero employment establishments. The original frame consists of the 225,977 units from 2007 that had zero employment, but have positive employment and positive monthly wages in 2010. Once the units in the frame of zero employment establishments were determined, average monthly wages, by aggregate industry, was calculated using 2010 employment and monthly wages. The frame average monthly wage (AMW) was calculated as follows:

$$AMW_{Frame} = \frac{\sum_{Cell} Employment * MonthlyWages}{\sum_{Cell} Employment}$$

For the simulated samples, average monthly wages, by aggregate industry, were calculated using employment from 2008, when the unit sampled in 2007 would have been collected, and monthly wages from 2010. The sampling weight is a result of the sampling process which was affected by the alternative

measure of size. Calculating the sample average monthly wage among sampled zero employment establishments for each of the alternative measures of establishment size was done as follows:

$$AMW_{Sample} = \frac{\sum_{Cell} Employment * SamplingWeight * MonthlyWages}{\sum_{Cell} Employment * SamplingWeight}$$

Given that the number of zero employment establishments sampled ranged in average employment from 91 and 685, and depending on the measure of size used, the number of units used in estimation is relatively few – between 11 and 85. Therefore, estimates from sample to sample vary resulting in volatile standard deviations when comparing across 200 samples. The results are provided in Tables 3 and 4 below.

Table 3: Average Sample Estimates of the ZEE Average Monthly Wage / ZEE Frame Estimate

	MOS_1	MOS_4	MOS_8	MOS_AVE (1)	MOS_AVE (4)	MOS_AVE (8)	MOS_MAX (1)	MOS_MAX (4)	MOS_MAX (8)
Overall	73.7%	73.6%	74.6%	72.3%	71.3%	78.6%	69.5%	74.8%	74.1%
Education	112.9%	165.0%	115.3%	82.5%	117.6%	109.0%	110.3%	104.5%	107.7%
Finance	98.9%	91.3%	87.4%	111.3%	88.6%	91.9%	79.1%	83.6%	80.2%
Goods Producing	79.5%	74.2%	81.3%	93.8%	82.7%	84.0%	79.7%	76.8%	86.8%
Health Care	199.5%	127.8%	108.3%	110.6%	148.7%	100.2%	146.5%	135.5%	131.8%
Service Producing	102.0%	96.7%	98.7%	86.7%	86.4%	104.5%	81.7%	93.0%	90.0%

Table 4: Relative Standard Error for Average Monthly Wages

	MOS_1	MOS_4	MOS_8	MOS_AVE (1)	MOS_AVE (4)	MOS_AVE (8)	MOS_MAX (1)	MOS_MAX (4)	MOS_MAX (8)
Overall	77.5%	52.9%	53.1%	74.0%	69.5%	91.2%	59.3%	56.3%	56.1%
Education	189.8%	491.8%	129.1%	111.0%	205.6%	125.8%	131.1%	132.0%	161.7%
Finance	193.0%	115.5%	90.9%	182.2%	140.8%	187.6%	97.9%	107.3%	100.7%
Goods Producing	146.1%	47.8%	56.6%	171.5%	70.3%	50.8%	72.7%	46.8%	84.5%
Health Care	220.9%	130.5%	109.8%	132.3%	179.4%	91.8%	204.3%	167.6%	227.4%
Service Producing	83.4%	90.5%	74.4%	98.8%	94.4%	139.0%	77.2%	74.1%	71.2%

If the sample reflects the frame well, the percentages in Table 3 will be near 100%. The best-performing measure of size varies depending on the industry. Some measures of size predicted the zero employment establishment average monthly wages for different aggregate industries better than others, suggesting that using industry trends may help assign an appropriate measure of size. MOS_AVE (8) performed the best at predicting average monthly wages. However, Table 4 shows that all measures of size had estimates with large relative standard errors, mostly due to small sample sizes that ranged in average employment from 11 and 85 establishments. The relative standard errors do not suggest that any alternative measure of size performs better than another.

Running estimation for the entire sample, 9,754 units, gives an indication of the potential impact of zero employment establishments on the overall estimates. These estimates more accurately mirrored actual NCS production procedures by adding benchmarking factors, though non-response factors were still not included in the calculation. For the frame, average monthly wages were calculated using employment and monthly wages from 2010. The sample estimates used employment from 2008, when the unit sampled in 2007 would have been collected, and monthly wages from 2010. The formulas are as follows:

$$AMW_{Frame} = \frac{\sum_{Cell} Employment * MonthlyWages * Benchmark}{\sum_{Cell} Employment * Benchmark}$$

$$AMW_{Sample} = \frac{\sum_{Cell} Employment * SamplingWeight * MonthlyWages * Benchmark}{\sum_{Cell} Employment * SamplingWeight * Benchmark}$$

The following tables illustrate the performance of each measure of size over 200 samples, when compared with the frame estimate for average monthly wage.

Table 5: Average Full Sample Estimates of the Average Monthly Wages / Frame Estimate

	MOS_1	MOS_4	MOS_8	MOS_AVE (1)	MOS_AVE (4)	MOS_AVE (8)	MOS_MAX (1)	MOS_MAX (4)	MOS_MAX (8)
Overall	95.8%	95.5%	95.5%	95.7%	95.3%	95.4%	95.7%	95.7%	95.6%
Education	100.4%	100.2%	99.7%	100.6%	100.5%	100.4%	100.1%	100.3%	100.2%
Finance	91.9%	90.9%	92.0%	92.3%	91.9%	91.9%	91.4%	91.8%	91.9%
Goods Producing	94.8%	94.7%	94.8%	95.1%	97.6%	95.2%	94.9%	95.0%	94.9%
Health Care	97.5%	97.5%	97.5%	97.1%	97.0%	97.4%	97.3%	98.1%	97.6%
Service Producing	95.3%	94.9%	94.7%	94.8%	95.1%	94.7%	94.8%	95.1%	94.9%

Table 6: Relative Standard Error for Average Monthly Wages

	MOS_1	MOS_4	MOS_8	MOS_AVE (1)	MOS_AVE (4)	MOS_AVE (8)	MOS_MAX (1)	MOS_MAX (4)	MOS_MAX (8)
Overall	3.2%	2.7%	2.1%	2.5%	2.4%	2.3%	2.4%	2.6%	2.2%
Education	4.4%	3.8%	2.8%	3.6%	4.6%	5.0%	3.6%	4.0%	3.7%
Finance	6.5%	6.0%	5.5%	7.0%	6.7%	5.5%	4.9%	6.0%	6.7%
Goods Producing	2.4%	2.6%	2.6%	3.0%	11.3%	2.7%	2.4%	2.4%	2.4%
Health Care	5.7%	6.6%	4.0%	4.4%	4.8%	4.3%	4.3%	4.5%	5.0%
Service Producing	5.7%	4.6%	3.4%	3.1%	3.0%	3.1%	3.1%	4.7%	4.0%

Using the full sample, both estimates and relative standard errors are much improved. There seems to be a trend of underestimating four of the five industries, regardless of the measure of size used. All of the alternative measures of establishment size performed well at predicting the average monthly wages for the frame. However, Table 6 shows that the relative standard errors are slightly better for the following alternative measures of size: MOS_MAX (1), MOS_AVE (8), and MOS_8.

Some establishment surveys allow zero employment establishments to have no chance of selection. Considering this condition, 200 additional simulations were run where each newly forming unit was given zero chance of selection and all other establishments on the frame were given a measure of size equal to the employment in the reference month. Seasonal units were given a measure of size equal to 1, 4, 8, the average, and the maximum. The following tables show the results of the process under the condition that no newly forming establishments are given a chance of selection.

Table 7: Average Estimates without Newly Forming Establishments in the Sampling Frame

	MOS_1	MOS_4	MOS_8	MOS_AVE	MOS_MAX
Overall	95.7%	95.7%	95.9%	95.6%	95.4%
Education	100.3%	99.7%	99.9%	100.2%	100.0%
Finance	92.0%	92.0%	91.4%	92.3%	91.6%
Goods Producing	95.3%	94.8%	95.7%	94.8%	95.2%
Health Care	97.3%	97.7%	97.8%	97.9%	96.9%
Service Producing	94.9%	95.0%	95.1%	94.7%	94.6%

Table 8: Relative Standard Error for Average Monthly Wages

	MOS_1	MOS_4	MOS_8	MOS_AVE	MOS_MAX
Overall	2.4%	2.5%	2.8%	2.0%	2.3%
Education	4.3%	4.3%	5.3%	4.1%	3.0%
Finance	5.7%	6.6%	7.1%	5.9%	5.4%
Goods Producing	5.3%	2.4%	8.3%	2.4%	2.5%
Health Care	4.7%	6.2%	6.9%	5.8%	4.3%
Service Producing	3.5%	4.3%	3.7%	3.3%	3.9%

The results in Table 7 show that, similar to using the full sample and including newly forming units, all measures of size were able to estimate the average monthly wage of the frame well. However, using the maximum as the measures of size seemed to perform the best when considering the relative standard errors (Table 8) of the average monthly wage estimates.

7. Conclusion and Future Research

Establishments with zero employment on the frame pose a problem for sampling, estimation, and collection. Ideally, any unit with employees would have an appropriate chance of selection and a sample would accurately represent all units in the frame as a result of correct sampling weights being assigned during sample selection. The goal of this research was to improve the assignment of establishment measure of size for units on the frame with zero employment.

First, an analysis of the frame was done to discover prevalence and trends related to establishments with zero employment in the reference month. After a study of the frame, simulated samples were drawn using new NCS procedures. Sample and estimation results for units with zero employment were analyzed for each alternative measure of size used.

The frame data showed that about 10% of the newly forming establishments could be expected to have a positive employment and be eligible to contribute to estimation two to three years later. These units are likely to have an employment of one or two, though one percent of them employ more than 9 workers. Since newly forming units have an average and maximum equal to zero, a fixed value must be assigned as the measure of size. Seasonal establishments on the frame tend to be eligible to contribute about 31% of the time. Since these units have some months of employment over the span of one year, using an average or maximum to set the measure of size takes advantage of employment data that is already known.

Zero employment establishments with the measure of size set to high fixed values were sampled more often. This resulted in more units appearing in the sample, but also many more units would be considered lost at the time of collection. Using a fixed value of one resulted in the highest percentage of units eligible to contribute to estimation by 2010. Not surprisingly, setting the measures of size to the maximum resulted in sampling more units than when the average was used.

For estimation, different measures of size performed better for different industries. Using the full sample, all measures of size performed equally well when comparing sample estimates to the frame estimates. Some measures of size had slightly lower relative standard errors. When sampling without giving newly forming units a chance of selection, the results were nearly the same as using the fully representative sample.

Perhaps the best choice of measure of size depends on the combination of low standard errors and a small percentage of zero employment establishments lost in collection. An ideal measure of size would account for trends among industries and types of units, approaching newly forming units, seasonal units, and consistently contributing units differently.

The following is a list of potential future research projects that are being considered in order to ensure that the best MOS is used in the selection of the NCS sample:

- Investigate trends among seasonal establishments that are not newly forming units,
- Investigate seasonal establishments that are zero employment units,
- Investigate seasonal establishments that have positive employment in the reference month,
- Consider the use of different reference months,
- Search for employment trends at the industry level for newly forming units and seasonal establishments,
- Define and test a custom measure of establishment size that accounts for employment trends within industry, and
- Investigate the current loss rates for zero employment establishments.

References

Ferguson, Gwyn R., Ponikowski, Chester, and Coleman, Joan (2010), "Evaluating Sample Design Issues in the National Compensation Survey", *2010 Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association, http://www.bls.gov/osmr/abstract/st/st100220.htm.

Ferguson, Gwyn R., Coleman, Joan, Ponikowski, Chester H. (2011), "Update on the Evaluation of Sample Design Issues in the National Compensation Survey", *2011 Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association. http://www.bls.gov/osmr/abstract/st/st110230.htm.

U.S. Bureau of Labor Statistics (2008) *BLS Handbook of Methods*, National Compensation Measures, Chapter 5. http://www.bls.gov/opub/hom/pdf/homch5.pdf

U.S. Bureau of Labor Statistics (2008) *BLS Handbook of Methods*, National Compensation Measures, Chapter 8. http://www.bls.gov/opub/hom/pdf/homch8.pdf

Note: Any opinions expressed in this paper are those of the author(s) and do not constitute policy of the Bureau of Labor Statistics.

Appendix A: Areas, Aggregate Industries, and Detailed Industries

<u>Areas</u>

Atlanta-Sandy Springs-Marietta, GA Metro
Boston-Worcester-Manchester, MA-NH CSA
Chicago-Naperville-Michigan City, IL-IN-WI CSA
Dallas-Fort Worth, TX CSA
Detroit-Warren-Flint, MI CSA
Houston-Baytown-Huntsville, TX CSA
Los Angeles-Long Beach-Riverside, CA CSA
Minneapolis-St. Paul-St. Cloud, MN-WI CSA
New York-Newark-Bridgeport, NY-NJ-CT-PA CSA
Philadelphia-Camden-Vineland, PA-NJ-DE-MD CSA
San Jose-San Francisco-Oakland, CA CSA
Seattle-Tacoma-Olympia, WA CSA
Washington-Baltimore-Northern Virginia, DC-MD-VA-WV CSA
Miami-Fort Lauderdale-Miami Beach, FL
Phoenix-Mesa-Scottsdale, AZ
Rest of New England
Rest of Middle Atlantic
Rest of East South Central
Rest of South Atlantic
Rest of East North Central
Rest of West North Central
Rest of West South Central
Rest of Mountain
Rest of Pacific

Aggregate and Detailed Industries

Aggregate Industry	Detailed Industry	NAICS Lower Bound	NAICS Upper Bound
Education	Educational Services (Rest of)	610000	611099
Education	Educational Services (Rest of)	611400	619999
Education	Elementary and Secondary Scho0ols	611100	611199
Education	Junior Colleges, Colleges and Universities	611200	611399
Finance, Insurance and Real Estate	Finance (Rest of)	520000	523999
Finance, Insurance and Real Estate	Finance (Rest of)	525000	529999
Finance, Insurance and Real Estate	Insurance	524000	524999
Finance, Insurance and Real Estate	Real Estate, Renting, Leasing	530000	539999
Goods Producing	Mining	210000	219999
Goods Producing	Construction	230000	239999
Goods Producing	Manufacturing	310000	336410
Goods Producing	Manufacturing	336412	339999
Health Care, including Hospitals and Nursing Care	Healthcare, Social Assistance (Rest of)	620000	621999
Health Care, including Hospitals and Nursing Care	Healthcare, Social Assistance (Rest of)	624000	629999
Health Care, including Hospitals and Nursing Care	Hospitals	622000	622999
Health Care, including Hospitals and Nursing Care	Nursing and Residential Care Facilities	623000	623999
Service Providing	Utilities	220000	229999
Service Providing	Wholesale Trade	420000	429999
Service Providing	Retail Trade	440000	459999
Service Providing	Transportation and Warehousing	480000	499999
Service Providing	Information	510000	519999
Service Providing	Professional, Scientific, Technical	540000	549999
Service Providing	Management of Companies and Enterprises	550000	559999
Service Providing	Admin., Support, Waste Management	560000	569999
Service Providing	Arts, Entertainment, Recreation	710000	719999
Service Providing	Accommodation and Food Services	720000	729999
Service Providing	Other Services (excl Public Administration)	810000	813999
Service Providing	Other Services (excl Public Administration)	815000	819999