Maximizing Sample Overlap Between Two Independent Surveys

David Piccone, Shail Butani, and Edwin Robison
Bureau of Labor Statistics
2 Massachusetts Ave NE, Suite 4985, Washington DC, 20212

Abstract
In 2010, the U.S. Bureau of Labor Statistics[1] (BLS) began preparations to embark on a "green jobs" initiative. The goals are to provide information for the U.S. economy on: 1) number of green jobs and trends over time; 2) industrial, occupational, and geographic distribution of these jobs; and 3) what these jobs pay. A new Green Goods and Services (GGS) Survey of 120,000 units is designed to measure the number of green jobs and trends over time as well as industrial and geographical distribution; the data collection began in 2011. The occupational distribution and what these jobs pay are being measured through BLS existing Occupational Employment Statistics (OES) Survey of 1.2 million units. For various reasons, both GGS and OES have independent sample designs--allocation, selection, data collection, and estimation. A major goal of this project is to maximize sample overlap between GGS and OES in order to reduce the cost of data collection. In this paper, we discuss natural sample overlap, a swapping algorithm used to force overlap of nearly identical sample units, and a sub-sampling approach.

Keywords: Power allocation, Neymann allocation, sample rotation panels, independent sample designs, sub-sampling, and swapping algorithm.

1. Introduction

In spring of 2011, the U.S. Bureau of Labor Statistics (BLS) began collecting data on employment related to production of green goods and services using the Green Goods and Services (GGS) survey. The GGS is a new Bureau survey that will collect data on the share of revenue or employment associated with production of green goods or services for business establishments in the 50 US States and the District of Columbia. For the purpose of this paper we will be referring to this employment as green employment. The BLS will also expand the existing Occupational Employment Statistics (OES) survey to collect occupational employment and wage data for establishments included in the GGS survey. This will allow BLS to produce occupational employment and wage estimates for businesses that report having green employment vs. businesses that report not having green employment.

A 100 percent sample overlap between the GGS and OES would be the most ideal situation for producing these types of estimates. This would allow information on green employment to be collected by the GGS survey, and occupational employment and wage information to be collected by the OES survey for every establishment in the GGS sample. The only way to achieve the 100 percent sample overlap between these two surveys would be to select GGS as a subsample of the OES sample. For several different reasons, which we will mention later in the paper, selecting GGS as a subsample was not a statistically viable option.

[1] Views expressed in this paper are those of the authors and do not necessarily reflect the views of policies of the Bureau of Labor Statistics

This paper describes methods researched and implemented for producing green employment and occupational estimates, while keeping the GGS and OES sample designs independent. Sections 2 and 3 provide brief descriptions of the GGS and OES sample designs, respectively. Section 4 describes the shortcomings of using the OES sample as the GGS sampling frame. The methods used to maximize the sample overlap between the two surveys and how to collect occupational data for the non-overlapping GGS sample are described in Section 5. Lastly, conclusions and options for future research are outlined in Section 6.

2. Description of the GGS Sample Design

The first time data will be collected for the GGS survey is May of 2011, or the second quarter of 2011 (2011Q2). There were no historical employment data associated with green goods or services available to help with the initial sample design. The sample design ensures a minimum reliability for the two main GGS estimation domains – state by major industry sector and national by detailed industry. The GGS uses the North American Industry Classification System (NAICS) for industry definitions.

2.1 GGS Frame Creation

The GGS sampling frame is a subset of all business establishments in the 50 U.S. States and the District of Columbia. Out of the 1,192 detailed 6-digit NAICS industries, 333 have been identified to be of specific interest and are in-scope for the GGS survey (Viegas 2011). These industries were thought to be the most environmentally friendly and would contain the majority of the green employment. Private and Government (Federal, State, and Local) establishments are included on the frame, excluding any establishment with an average employment of zero over the past 12 months.

The GGS uses the BLS Quarterly Census of Employment and Wages (QCEW) as its sampling frame. The data for the QCEW comes from State Unemployment Insurance files that are collected by individual State agencies. These files are made up of several descriptive variables such as name, address, monthly employment counts, industry classification, and geography information for nearly all establishments in the United States. It takes about one year for these data to be processed, meaning the GGS frame for the 2011Q2 sample is comprised of 2010Q2 QCEW data. The 2010Q2 QCEW has over 8 million business establishments accounting for about 150 million employees. The GGS sample frame is restricted to the 333 in-scope industries and has approximately 1.8 million establishments accounting for about 30 million employees.

About 13,000 in-scope establishments comprising of about one million employees were pre-identified as being involved with some kind of green activity. These units were identified internally by BLS by use of the internet and an environmental database maintained by Environmental Business International (an environmental publishing, research and consulting company). In this paper these 13,000 establishments will be referred to as the environmental establishment frame. These establishments will have special treatment during the GGS allocation and selection phases.

The 2010Q2 QCEW covered a large number of intermittent employees hired for the 2010 Decennial Census. Since almost all of these employees will no longer be working at the time GGS will collect its data (May of 2011) the establishments with these employees were deemed out-of-scope for the GGS survey.

2.2 GGS Sample Allocation

The GGS has funding for a sample size of about 120,000 establishments, where 116,000 establishments will be selected in a second quarter initial sample and 4,000 will be selected in a fourth quarter birth sample. The initial sample is divided in the following way:

Table 1: GGS Allocation Breakout

Type of Frame Unit	Sample Allocated
Private Establishments	94,500
Local Government Establishments	7,700
State Government Establishments	4,000
Federal Government Units	3,300
Environmental Establishments	6,500
Total:	116,000

Each type of frame units has its own independent allocation. We will briefly explain each allocation below.

2.2.1 The GGS Private Establishment Allocation

The GGS private establishment allocation can be thought of as two separate allocations, one that stratifies the frame by state / 2-digit NAICS industries and the other that stratifies by 4 or 6 digit NAICS industries. These 4 or 6 digit industries will be called Allocation NAICS, or A_NAICS for the remainder of this paper. For the most part the A_NAICS industries are at the 4-digit NAICS detail, however some industries that seemed to be highly environmental (ex. 221119 – Other Electric Power Generation) were held out to the 6-digit detail.

The GGS private sample is first allocated by giving a minimum of 40 sample units to each state by 2-digit NAICS stratum. If there are less than 60 establishments within a stratum they all are allocated into the sample. This allocated about 24,000 sample units for the 2011Q2 sample. Next 1,000 sample units are allocated within each state using a power allocation (Bankier, 1988):

$$n_h = n_s \frac{\sqrt{X_h}}{\sum_{h \in s} \sqrt{X_h}} \qquad (2.1)$$

Where,

$n_h =$ the amount of sample allocated to stratum h (state by 2-digit NAICS)

$n_s =$ the state sample size, which is 1,000

$X_h =$ the number of employees in stratum h

After the minimum and power allocation, about 60,000 sample units were allocated for private establishments in 2011Q2. Thus, a sample of 60,000 ensures a minimum sample of 1,000 per state and 40 establishments for each 2-digit NAICS within a state.

Next sample is allocated nationally to A_NAICS industry strata, using the following power allocation:

$$n_h = n\frac{\sqrt{X_h}}{\sum_{all\ h}\sqrt{X_h}} \qquad (2.2)$$

Where,

n_h = the amount of sample allocated to stratum h (A_NAICS)

n = the national sample size

X_h = the number of employees in stratum h

The national sample size is iteratively increased until the total private allocation, after reconciling the state by 2-digit and national A_NAICS allocations, is close to 94,500. The last step of the private allocation is to set a minimum of 40 sample units to each 6-digit NAICS stratum. If there are less than 60 establishments within a stratum they are all allocated into the sample. In the 2011Q2 sample there were a total of 94,800 sample units allocated for the private sample.

2.2.2 The Local, State and Federal GGS Allocations

The sample units for Local, State, and Federal establishments are allocated the same way. The frame is stratified into state and 2-digit NAICS industry strata and a minimum allocation is used. A minimum sample of 40 units is allocated to each state by 2-digit NAICS stratum. If there are less than 60 establishments within a stratum they are all allocated into the sample. In the 2011Q2 sample the Local, State, and Federal samples were allocated, respectively, about 3,000, 3950 and 7,700 sample units.

2.2.3 Environmental GGS Allocation

The environmental allocation includes establishment in the private and government sectors. The frame is stratified by 6-digit NAICS industry and size class. Size classes are seven categories that put establishments of similar size together. For example, if an establishment has 1 to 9 employees if would be in size class 1 for GGS. The environmental sample is allocated using the following rules:

$$n_h = \begin{cases} N_h/_5 & for\ size\ class\ 1 \\ N_h/_3 & for\ size\ class\ 2 \\ N_h/_2 & for\ size\ class\ 3 \\ N_h & for\ size\ class\ 4+ \end{cases} \qquad (2.3)$$

Where,

n_h = the amount of sample allocated to stratum h (6-digit NAICS by Size Class)

N_h = the number of frame units in stratum h

In 2011Q2 there were about 6,550 sample units allocated for the environment sample

2.3 GGS Sample Selection

The Private and Government samples are selected using a probability proportionate to size where the size for an establishment is defined below:

$$size_i = \begin{cases} 10 & if\ X_i \leq 10 \\ X_i & if\ X_i > 10 \end{cases} \quad (2.4)$$

Where,

$size_i$ = unit i's measure of size

X_i = unit i's max employment

This type of sampling is sometimes referred to as PPZ sampling (Cochran 1977). The smallest establishments are treated differently because of the assumption that they have the potential for very large relative employment shifts between the time period of the QCEW data on the frame and when the establishment is sampled. By raising the size of the smallest establishments the selection probabilities are raised causing the weights to be lower and more stable. If GGS was selected using a straight PPS sampling approach there is a potential for the smallest units to have very large weights which would then be multiplied by a high employment number if there was a big shift in employment.

The environmental sample is selected using simple random sampling within each 6-digit NAICS by size class stratum. Since the sample is allocated at a higher rate as the size class increases, there is an implicit probability proportionate to size selection scheme.

3. Description of the OES Sample Design

The OES survey is designed to collect occupational employment and wage data on employees working in the 50 states, the District of Columbia, the Virgin Islands, Puerto Rico and Guam. The main estimation domain is at the detailed Metropolitan Statistical Areas (MSA) and residual areas within each state that are called Balance of State (BOS) areas. In order to produce estimates at such detail, a sample of 1.2 million business establishments is selected over three years in bi-yearly samples. A sample of 200,000 establishments is selected in the second and forth quarter of each year (BLS Handbook of Methods 2011).

3.1 OES Frame Creation

The OES survey also uses the QCEW as its sampling frame. The majority of the 1,192 NAICS industries are in scope for OES except for most of the agriculture sector (except Logging NAICS 113310, support activities for crop production NAICS 1151, and support activities for animal production NAICS 1152). Private household (NAICS 814) are also excluded (BLS Handbook of Methods 2011). The 2011Q2 OES frame had about 7 million in-scope business establishments which account for about 150 million employees.

3.2 OES Sample Allocation

The OES frame stratification is by state, MSA or BOS area, and 4 or 5 digit NAICS industries. The majority of the strata use 4-digit NAICS detail but some industries have unique occupational distributions at the 5-digit detail which are stratified at more detail. These 4 or 5 digit NAICS industries will be referred to as allocation NAICS, or A_NAICS.

For each bi-yearly sample a full 1.2 million establishment sample is first allocated, and then the allocation is divided by six at the stratum level. First, a minimum sample allocates the sample using the following rules:

$$n_h = \begin{cases} N_h & if\ N_h \leq 3 \\ 3 & if\ 4 \leq N_h \leq 11 \\ 6 & if\ N_h \geq 12 \end{cases} \quad (3.1)$$

Where,

n_h = the amount of sample allocated to stratum h (State by MSA/BOS by A_NAICS)

N_h = the number of frame units in stratum h

Next, the sample is allocated using a power Neyman allocation, using the following formula (Lawley 2007):

$$n_h = n\frac{\sqrt{X_h}\,S_h}{\sum_{all\,h}\left(\sqrt{X_h}\,S_h\right)} \quad (3.2)$$

Where,

n_h = the amount of sample allocated to stratum h (State by MSA/BOS by A_NAICS)

n = the national sample size

X_h = the number of employees in stratum h

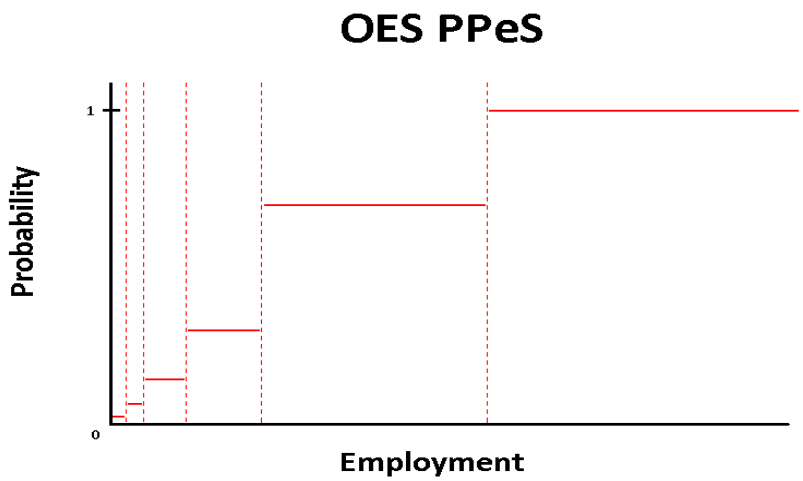
S_h = the measure of occupational employment variability within stratum h

The final amount of sample allocated for each stratum is the maximum of the minimum and power Neyman allocations. The national sample size used in formula 3.2 is iteratively changed until the final amount of sample allocated, after reconciling the two different allocations, is about 1.2 million. The last step of the OES allocation is to divide each stratum allocation amount by six, to get the final allocation for the bi-yearly sample.

3.3 OES Sample Selection

After the sample is allocated, the bi-yearly sample is selected using a probability proportionate to size approach. Every establishment within an OES-defined size class is given the average employment value for that size class. This is a step-wise probability proportionate to size scheme, which is another version of a PPZ sampling approach (Cochran 1977). Below is visual representation of the OES sample selection scheme:

Image 1: OES Sample Selection Approach



4. Issues with Using the OES Sample as the GGS Frame

During the early stages of the GGS sample design research, we explored the idea of selecting the GGS sample as a sub-sample of the OES survey. This would achieve a 100 percent sample overlap between the two surveys. We found that this would not be a statistically viable option for the following reasons.

The main issue with using the OES sample as a sampling frame for GGS is the combined OES sample is not representative of the most current time period. The full OES sample is the combination of six different samples selected over the past three years, and does not represent any one time period but rather an average over the six time periods. This lack of representativeness would lead to an inefficient GGS sub-sample causing higher variances for the estimates.

Another issue is the OES full sample has many establishments that are currently out-of-business that were sampled off of the previous five OES frames. About 7.2 percent of the 2011Q2 full OES sample (combination of 2011Q2, 2010Q4, 2010Q2, 2009Q4, 2009Q2 and 2008Q4 samples) was no longer in-business. This is not a problem for OES estimates since its sample was designed to represent a pseudo three year average frame, however the GGS sample is designed to represent the most current year and if selected as a sub-sample would introduce bias to the GGS estimates. At the same time, the younger or new units are underrepresented for GGS purposes. For these reasons we determined that the GGS could not be selected as a sub-sample of OES, and began looking at different options for collecting occupational data for the GGS sample.

5. Methods Used to Collect Occupational Data for GGS

After determining that a sub-sample approach was not feasible, we decided to keep the OES and GGS sample designs as independent as possible. The benefit of independent sample designs is that any sample design changes or issues would be confined to only the one survey instead of both. This was an important research goal to have since sample designs can change often due to budget constraints, changes to stakeholder's needs, changes to standardized classification systems, or improvements to the overall methodology. We decided on selecting the GGS and OES samples using completely independent sample designs, and then using an algorithm to increase the overlap by

replacing non-overlapping GGS sampled units with non-overlapping OES units that had similar characteristics. Lastly we would draw a sub-sample of the still non-overlapping GGS sampled units and collect occupation employment and wage data for these units. This approach is a statistically defensible option for collecting green employment and occupational employment for the GGS sample.

5.1 Natural Sample Overlap between GGS and OES

In both the GGS and OES sample designs, a greater probability of selection is given to establishments with more employees. This causes a substantial amount of overlap between the two samples, even though they have independent sample designs. In 2011Q2 41 percent (about 41,300 sampled establishments) of the GGS sample overlaps naturally with the OES sample. The overlap is higher for large establishments, decreasing as establishments get smaller. This causes the sample employment overlap to be significantly larger than the unit overlap, at 80 percent (about 8.3 million employees).

Most of the State and Local Government units had to be excluded from these sample overlap counts because OES and GGS define their public Primary Sampling Units (PSUs) differently. In the OES sample, State and Local government PSUs are aggregated to specific geographic areas to make data collection easier for the state data collectors. In the GGS sample, State and Local government PSUs are single business establishments. Only OES State and Local aggregate PSUs that contain only one establishment are used when identifying the natural overlap between the two surveys and in the replacement algorithm described in section 5.3. All Federal Units are also excluded because OES currently receives a census of Federal data from the US Office of Personnel Management (OPM) and can possibly link units sampled for the GGS survey back to this data.

In Table 2 below the amount of sample unit and employment overlap is summarized by GGS-defined size classes is shown:

Table 2: GGS and OES Natural Overlap by Size Class

		Sample Unit Overlap			Sample Employment Overlap		
Size Class	Empl Range	Total GGS Sample Units	Total OES Overlap	Percent Overlap	Total GGS Sample Empl	Total OES Overlap	Percent Overlap
1	1 to 9	35,959	5,335	15%	130,020	24,243	19%
2	10 to 19	11,673	3,818	33%	164,554	54,601	33%
3	20 to 49	18,641	8,384	45%	604,296	274,890	45%
4	50 to 99	13,544	7,595	56%	960,263	539,347	56%
5	100 to 249	12,189	7,666	63%	1,891,813	1,201,563	64%
6	250 to 499	4,908	4,468	91%	1,698,948	1,548,145	91%
7	500 +	4,218	3,998	95%	4,880,156	4,655,885	95%
	Totals:	101,132	41,264	41%	10,330,050	8,298,674	80%

Table 2 shows how the sample overlap is skewed towards the larger establishments selected for each survey.

5.2 Replacement Algorithm

To increase the overlap for the smaller establishments we used an algorithm that replaces non-overlapping GGS sampled units with non-overlapping OES sampled units. All establishments from the environmental frame were excluded from this process since they were pre-determined as having green activity and important to the GGS sample. We used strict replacement criteria to minimize any bias this process could introduce. In order for an establishment sampled for GGS to be replaced by one from OES it must meet the following criteria:

- Industrial Criterion: The GGS and OES sampled establishments must have the same 6-digit NAICS industry classification
- Geographic Criterion: The GGS and OES sampled establishments must be within the same state, first giving preference to establishments in the same State and MSA/BOS area, then relaxing the search to just State
- Age Criterion: The GGS and OES sampled establishments must have begun their business in the same year and quarter if they have been in business less than three years
- Employment Criterion: The GGS and OES sampled establishments much meet the following employment tolerances:

GGS Employment	OES Employment
1 to 4	1 to 4
5 +	$0.9 * GGS\_Emp$ to $1.1 * GGS\_Emp$

- Multi-Establishment Criterion: The GGS and OES sampled establishments must both be from a company that has multiple establishments, or both from a company that has only one establishment

While researching this replacement algorithm several different employment tolerances were tested. As the tolerance was relaxed there was a trade-off between the number of sampled units we were able to replace, with the amount of employment bias (the amount of employment brought into the GGS sample vs. the amount removed) introduced. This relationship can be seen in Table 3.

Table 3: Amount of Sample replaced and Employment bias at different Employment Tolerances

Tolerance	# of replace-ments	% of non-overlapped GGS	GGS Empl Removed	OES Empl Added	Bias	%
0.2	24,876	41.6%	425,865	434,726	8,861	2.1%
0.1	21,894	36.6%	344,791	349,564	4,773	1.4%
0.05	19,843	33.1%	266,686	269,933	3,247	1.2%
0.04	19,316	32.3%	243,336	246,046	2,710	1.1%
0.03	18,748	31.3%	214,746	216,945	2,199	1.0%
0.02	18,179	30.4%	182,226	184,009	1,783	1.0%
0.01	17,751	29.7%	148,529	149,962	1,433	1.0%
0	17,648	29.5%	134,153	135,521	1,368*	1.0%

* Please NOTE that the reason why there is still a Bias when using a zero percent employment tolerance is because we allow any GGS establishments with five employees or less to be replaced by any OES establishment with five employees or less.

We chose the ten percent employment tolerance because it gave a significantly better bias percentage than the twenty percent, and it was only slightly worse (0.4 percent) than the zero percent tolerance. The gain to the number of replacements was significantly more (about 4,000) when using the ten percent tolerance compared to the zero percent tolerance. Since the algorithm replaced mostly smaller establishment, we looked closely at the amount of weighted employment bias that was introduced since the smaller sampled units have the largest weights. In Table 4 below, we compared the total frame employment with the weighted sample employment before and after the swapping algorithm:

Table 4: Frame Employment vs. Weighted Sample Employment before and After Algorithm

Frame Employment	Weighted Sample Employment Pre-Algorithm	Percent Diff	Weighted Sample Employment Post-Algorithm	Percent Diff
30,274,690	30,158,229	**-0.38%**	30,175,090	**-0.33%**

Table 4 shows that the amount of weighted employment bias introduced by the replacement algorithm is very small. We looked at similar comparisons at different geographic and industrial levels and the weighted employment bias was negligible.

In 2011Q2 after using this replacement algorithm the amount of sample overlap between the GGS and OES surveys increased to 64 percent (about 64,700 sampled establishments). The amount of sample employment overlap increased slightly to 83 percent (about 8.6 million employees).

5.3 Sub-Sample of Non-Overlapping GGS Sample

To collect occupational employment and wage data for the piece of the GGS sample that does not overlap with OES, a sub-sample of 25,000 establishments is selected. These establishments are asked additional information about which occupations their employees work in and how much their wages are. As a precaution 2,000 sample units out of the 25,000 units for the sub-sample were saved for Federal data in case the GGS sampled units can not be retrieved from the census of OPM data that OES receives.

The non-overlapping GGS sample is stratified by 6-digt NAICS industries and the sub-sample is allocated using the follow formula:

$$n_h = n \cdot \frac{N_h \sqrt{S_h P_h}}{\sum_{all\,h} \left(N_h \sqrt{S_h P_h} \right)} \qquad (5.1)$$

Where,

$n_h =$ the number of sub-sample units allocated to industry h (6-digit NAICS)

$n =$ the total sub-sample size

$S_h =$ the measure of occupational employment variability within stratum h

$N_h =$ the number of non-overlapping GGS sample units within stratum h

$P_h =$ the non-overlap employment percentage for stratum h

This allocation is driven by the amount of non-overlapping GGS sample and the occupational employment variability for a particular 6-digit NAICS industry. These were believed to be important factors that when increased for a particular industry would warrant an increase in sample size. After collecting data with the GGS survey we can re-evaluate our decision to use formula 5.1 for the sub-sample allocation.

The first step of selecting the sub-sample is to identify the units within each 6-digit NAICS industry that will be contributing to the variance estimate the most, and select them with certainty. For each non-overlapping GGS sampled unit the amount of the GGS universe they represent is calculated using:

$$E_i = w_i X_i \qquad (5.2)$$

Where,

$E_i =$ establishment i's weighted employment

$w_i =$ establishment i's GGS sampling weight

$X_i =$ establishment i's employment

Next the average weighted employment is calculated for each 6-digit NAICS industry by:

$$\bar{E}_I = \frac{1}{n_I} \sum_{i \in I} E_i \qquad (5.3)$$

Where,

$\bar{E}_I =$ the amount of weighted employment each sub-sample units will represent on-average

If any unit's E_i is greater than or equal to $\bar{E}_I$, then it's selected into the sub-sample with certainty. This is an iterative process, where each time establishments are selected with certainty, $\bar{E}_I$ is re-calculated and compared to the remaining unit's E_i. Once there are no more units to select with certainty, the remaining units are selected within each industry using simple random sampling (SRS). The final weight that will be used for the occupational estimates for the GGS sampled units selected into the sub-sample is the product of their original GGS weight and the inverse of their sub-sampling selection probability.

6. Conclusions and Future Research

The BLS green jobs initiative caused the creation of the new Green Goods and Services survey. The goals of this survey are to collect revenue or employment data associated with green goods or services and occupational employment and wage distributions within business establishments in the United States. This paper explained the research we've done to coordinate the new GGS survey with the existing OES survey in order to meet these goals. We are able to identify the natural overlap, increase this overlap by using an algorithm that replaces non-overlapping GGS sample with non-overlapping OES sample, and represent the non-overlapping GGS sample by sub-sampling. Since no data exists on

green employment we had to approach this research empirically, borrowing many techniques used in other surveys. Once we have data on green employment we will be able to improve on the research we've outlined in this paper.

We plan to do research on the different response situations for the GGS survey. There are four different ways we can get responses from a sampled establishment: 1.) response for green employment questions and response for occupational questions, 2.) non-response for green employment questions and response for occupational questions, 3.) response for green employment questions and non-response for occupational questions, and 4.) non-response for green employment questions and non-response for occupational questions. We will work to better understand these situations, and find appropriate way to handle them to reduce non-response error in our estimates.

Acknowledgements
Dixie Sommers, Bureau of Labor Statistics
George Stamas, Bureau of Labor Statistics
Laurie Salmon, Bureau of Labor Statistics
David Byun, Bureau of Labor Statistics
Justin McIllece, Bureau of Labor Statistics

References

Bankier, Michael D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *American Statistician*, Vol. 42, pp. 174-177.

Cochran, William G. (1977). *Sampling Techniques*, 3rd ed. John Wiley and Sons, New York.

Lawley, Ernest, Stetser, Marie, and Valaitis, Eduardas. (2007) *Alternative Allocation Designs for a Highly Stratified Establishment Survey*. 2007 Joint Statistical Meetings.

U.S. Bureau of Labor Statistics (2011). Handbook of Methods, Washington, DC. <http://www.bls.gov/opub/hom/home.htm>

Viegas, Robert, Fairman, Kristin, Haughton, Donald, Clayton Rick. (2011) *Measuring Green Industry Employment: Developing a Definition of Green Goods and Services.*2011 Joint Statistical Meetings