Supporting Statement for Request for OMB Approval
Quarterly Census of Employment and Wages Green Goods and Services
Data Collection Clearance

B. Collection of Information Employing Statistical Methods

The collection of data on Green Goods and Services (GGS) is new to BLS. No survey-specific historical information is available on variances, response rates, etc.

1a. Universe

Geographic coverage includes the 50 States and the District of Columbia. Private establishments and government units are included, but units with average employment of zero over the last 12 months are excluded. Data are to be collected for establishments in 333 detailed industries identified to be of specific interest for the GGS Survey. The industries are defined using the 6-digit detail of the North American Industry Classification System (NAICS; includes 1,193 6-digit industries). Attached are two tables summarizing the survey frame at both the industry sector level and at the State level for Private, Federal government, State government, and local government ownership (Section B attachments 1 and 2).

The sampling frame is the Quarterly Census of Employment and Wages (QCEW) Longitudinal Database (LDB) maintained by the Bureau of Labor Statistics. The QCEW has over 9 million establishments and includes items such as business name, address, geographic coding, 6-digit NAICS industry code, current information on employment and wages, and some past economic information. About 1.8 million establishments with employment of 30 million are in the 333 in-scope industries.

For the purposes of GGS sample allocation, we aggregate 333 detailed industries into 171 groups of industries or "allocation" NAICS (ANAICS). For most in-scope industries, the ANAICS is the 4-digit NAICS and includes all in-scope NAICS-defined industries within the 4 digit. Within some 4-digit industries, ANAICS splits out specific 5- and 6-digit NAICS industries where we anticipate having a higher incidence of green activity. ANAICS 2- and 3-digit coding is the same as for NAICS, though restricted to GGS-eligible industries.

Industry sectors are also defined for use in allocation. Industry sectors are 2-digit ANAICS with two exceptions. The manufacturing sector combines three 2-digit codes. The trade sector combines retail trade and wholesale trade.

About 13,000 in-scope "Green Frame" establishments with one million employees were pre-identified as having "green" activity. A database of likely green establishments was developed internally by BLS by comparing the QCEW data to internet search results using keywords associated with green industries. BLS also contracted with Environmental Business International (an environmental publishing, research, and consulting company), to obtain their database of green establishments for comparison to the QCEW. By comparing the information obtained through these sources and comparing the NAICS codes of these establishments on the QCEW, Green Frame establishments were matched to the QCEW and a "green" indicator will be used to assist in oversampling "green" establishments.

BLS will coordinate sample selection for GGS with the Occupational Employment Statistics (OES) survey to enable a cost-effective analysis of green employment by occupation.

1b. Sample
BLS will select about 120,000 establishments per year from the QCEW. The following figures are based on a sample selected using the QCEW frame for the second quarter of 2010. From the 13,000 Green Frame establishments, the largest 3,250 in employment will be selected with certainty, and 3,250 noncertainty establishments will be selected. About 95,000 establishments will be allocated to the private sector (profit and nonprofit, excluding Green Frame) and about 40,000 of those will be certainties with large employment. An additional 4,000 establishments are reserved for the three quarterly births samples. The remainder of the sample will be allocated to government units. Approximate sample sizes are as follows: 3,000 for Federal government; 4,000 for State government and 7,500 for local government. Although the precision is not specified, the goal is to publish industry sector data in every State and to obtain useful data at the national level for each ANAICS. If possible within budget constraints, some additional information by 6-digit NAICS is also desired. Attached are two tables summarizing a full GGS sample allocation industry sector by ownership and State by ownership (Section B attachments 3 and 4). Green Frame establishments are not separated but are included in columns for the private, Federal government, State government, and local government sectors.

Panel Rotation – The annual sample of about 120,000 establishments will be divided into 3 panels of about 40,000 establishments each. Each year, one panel of 40,000 establishments will be added to the survey and another panel of 40,000 units dropped. Once the survey is established, every new panel will be surveyed three years in a row, and then dropped.

- Larger establishments are selected with certainty and will be included in every panel (data collected every year).

- All establishments in "small" ANAICS industries are included as certainty.

- All establishments in "small" State x industry sector cells are included as certainty.

- Some additional Green Frame establishments are designated as certainty, because they may produce a rare product or have other unique "green" activity.

- A quarterly sample of births is planned, using the same probabilities applicable to continuing establishments in the frame. About 4,000 birth establishments will be selected per year.

Stratification – The Green Frame will be stratified by 6-digit NAICS and size class (1-9, 10-19, 20-49, 50-99, and 100+ employees) and systematic samples selected in the noncertainty strata. Green Frame establishments can be of any ownership, are processed separately, and are excluded from the other portions of the frame. Federal government stratification is State by industry sector. State government stratification is State by industry sector. Local government stratification is State by industry sector for these sectors: utilities; transportation and warehousing; professional, scientific, and technical services; remediation services; educational services; arts, entertainment, and recreation; public administration (all other sectors combined to a residual category). For private establishments (excluding the Green Frame)

three levels of stratification are examined during sample allocation: 1) State x industry sector, 2) national ANAICS, and 3) national 6-digit NAICS. Further stratification by establishment size did not prove to be practical.

PPES Sampling Method (excluding the Green Frame) – Noncertainty establishments will be selected with known probabilities. A modified method of probability Proportional to Estimated Size (PPES) sampling will be used. The estimated size (x) for an establishment will be the maximum employment for the last 12 months of data available. Within a stratum under PPES, a noncertainty establishment with twice the size of another would be sampled with twice the probability. This type of sampling tends to be efficient when the outcome variable (y) is highly correlated with size, and is ideal when there is a fixed x=Cy relationship (where C is a positive constant). There are no GGS data for analysis, but an analysis limited to the changes over time in the size measures indicates that smaller establishments should be sampled with somewhat higher probabilities than under standard PPES. To accomplish this, a modified PPES method will be used where a minimum employment size of x=10 is used for the smallest establishments. (Further analysis may lead to changes in the specified minimum size.) A higher proportion of smaller establishments may specialize in "green" activity, but no adjustment will be made at this time since no data have been collected. Substantial modification to original PPES probabilities are also needed when the allocation and sampling are controlled in three dimensions: 1) State x industry sector, 2) national ANAICS, and 3) national 6-digit NAICS.

Power Allocation – Trial allocations to strata based on allocations proportional to stratum size resulted in samples that were much too large for the States and industries with the most employment. This was an analytical judgment based on projected utility of the data, not a decision based on an optimization formula. The resolution was to use power allocations based on the square root of stratum size. The resulting allocations is a compromise between the extremes of 1) allocating a nearly equal sample size to each stratum and 2) proportional allocation to strata which would be preferred if high-level data were of prime importance. BLS found during testing that some industries still had overly large samples; as a result the sample for each ANAICS is capped at about 1.5% or less of the total sample.

Government – Federal government, State government, and local government will be treated as separate sectors. Within each sector, only minimum State by industry sector sampling criteria will be set. There will be no national allocation for government data.

2a. Sample Design
GGS panels will have a probability-based sample aimed at satisfying data needs at both the State x industry sector level and the national ANAICS level. The basic sampling unit is an establishment. After the initial start-up year, new noncertainty units will be placed in a panel that is surveyed three consecutive years, then will be out of the survey for one or more years. A modified form of PPES sampling is used since data are expected to be highly correlated with establishment size measured in employment. Restricted to in-scope industries, establishment on the QCEW frame are separated into 5 mutually exclusive parts that are separately sampled. Approximate sample counts refer to a sample selected from the QCEW frame for quarter 2 of 2010.
- Green Frame;   sample 6,500 ;   stratification industry by size class
  (can have any ownership code)
- Federal Government;   sample 3,000;   stratification state by industry sector
- State Government;   sample 4,000;   stratification state by industry sector

- Local Government;   sample 7,500 ;   stratification state by industry sector
- Private;   sample 95,000 ;   complex stratification using state, industry

Additionally, about 4,000 units from the private sector are reserved for the three quarterly birth samples.

Green Frame Sampling – All establishments in 6-digit ANAICS industries with few establishments will be sampled with certainty. All establishments in the largest size classes covering in excess of 80% of employment will be selected with certainty (about 3,250 establishments). Sampling fractions are set for the three smaller size classes and systematic samples selected from each 6-digit NAICS by size class stratum. About 3,250 noncertainty establishments are selected from this frame. Thus, a total of 6,500 establishments across all ownerships are selected from the Green Frame.

Federal Government Allocation – All of the largest establishments will be certainty. In each State, select all establishments with certainty from an industry sector stratum with few establishments. Within each state, allocate 40 sample units to other industry sector strata.

State Government Allocation – All of the largest establishments will be certainty. In each State, select all establishments with certainty from an industry sector stratum with few establishments. Within each State, allocate 40 sample units to other industry sector strata.

Local Government Allocation – All of the largest establishments will be certainty. Some industry sectors are collapsed to a residual (reducing the required sample by about 1,000 establishments). In each State, select all establishments with certainty from a collapsed industry sector stratum with few establishments. Within each State, allocate 40 sample units to other industry sector strata, except allocate 24 sample units per state for the collapsed residual category.

Allocation Steps for Private Establishments (excluding Green Frame) – Removing the large certainties, there are n sample units to allocate for private establishments. The allocation has 5 basic steps.

- State Allocation – Allocate n_s establishments to each State's ANAICS strata using a square root power allocation (initially used n_s = 1,000). Assuming a net design effect of ½, a Coefficient of Variation (CV) of 10% is obtained for a hypothetical "green" jobs rate of 5%. In addition, ensure a minimum allocation of 40 sampled establishments to each State x industry sector. Establishments in the smallest State x industry sector strata, those with few establishments, will be made certainty (selected with probability 1.000). In addition, compute (in probability) the allocations across all States to selected ANAICS aggregations and cap the allocation to industries at the national level at 1.5% of the total sample for private industries.

- National Allocation – Allocate n_{us} establishments to ANAICS industries using a square root power allocation. (An n_{us} of about 80,000 was used initially.) The allocation is done without reference to the State allocation. The exact number to allocate is determined when the national allocation and State allocations are reconciled. Ensure a minimum allocation of 40 sampled establishments for each 6-digit NAICS (not ANAICS) and make the smallest industries certainty. Also in this step, ensure that each ANAICS is allocated less than 1.5% of the total sample allocated.

4

- Reconcile the State and National Allocations – Each establishment in the universe will have a known probability of selection based on the State allocation and modified PPES sampling. In either allocation less than one establishment may have been allocated to a cell, but that is not a problem. To reconcile the allocations, select the largest of each establishment's probabilities. The reconciled probabilities can be used to compute allocations (in probability) for State x industry sector cells, by national ANAICS cells, by national NAICS, or for any other desired level of aggregation.

- Adjust Sampling Parameters – Several iterations and minor adjustments of parameters is needed to control the overall private sample to the desired n establishments. The number of sample units for the private and government sectors will also be continually examined.

Method of Selecting the Sample with Known Probabilities – For a full private sample (no panels) of n noncertainty establishments, a sample can easily be selected using the probabilities p determined for each establishment. For establishment i with probability p_i generate a random number r_i between 0 and 1. If r_i is less than or equal to p_i then select the establishment for the sample. This is an extended form of Bernoulli sampling called Poisson sampling. Many cells at the ultimate detail (State x NAICS) will have no sample, but that is to be expected. The method was tested and a sample resulted that closely matched test allocations for State industry sectors and national ANAICS.

Panel Sampling – For panel sampling, a simple expedient is to divide the universe into 3 parts using permanent random numbers. That way, an establishment is always in the same 1/3 universe. Selection of a panel from any sub-universe proceeds using the probabilities p determined for each establishment. When a panel is rotated out of the GGS survey, it would be replaced by a panel selected from the same sub-universe.

Coordination with OES – The Occupational Employment Survey includes about 200,000 establishments in data collection every six months. Estimates are made based on 6 panels of data collected over a 3-year span. For many detailed industries and/or States, it was found that the existing OES sample was not large enough to allow a subsample of OES to be taken that was sufficient to meet the needs of GGS. The possibility was researched of augmenting OES to allow GGS subsampling but this was determined to be infeasible due to budgetary constraints and the extensive modification of OES systems that would be required. The following protocol was tested and will be used.

- Select a GGS sample from the QCEW frame independently of OES. Select two OES panels from the same frame.
- Match the GGS sample to the 6 most recent OES frames. This "natural overlap" includes about 1/3 of sampled GGS establishments. Larger establishments are heavily represented since both designs select larger establishments with greater probabilities than smaller establishments.
- "Swap" establishments out of the GGS sample and replace with nearly identical OES units. A "nearly identical" unit must be in the same 6-digit NAICS and also be a near-match in terms of employment, geographic location (State and MSA), and age of establishment. Matching/swapping is possible in the private sector, in the local government sector excluding multi-unit reporters, and for hospitals and educational institutions in the State government sector. About 1/3 of GGS sample establishments can be swapped in this manner, resulting in an overall overlap (natural overlap plus swapped units) of about 2/3.

- From the remaining unmatched GGS units, select a probability subsample that stays within budgetary constraints (about 20,000 establishments) for OES purposes.

Changes in the Sample Design – When GGS data become available, it will be possible to more efficiently design the sample to meet targeted data needs. It is anticipated that changes in industry scope may be made.

2b. Estimation Procedure
GGS estimators of total will take the form of a Horvitz-Thompson estimator. Let a cell have c sample units. Each establishment has a weight w_i that is the inverse of its probability of selection. Each establishment has a data value y_i. For example, the data value could be the total proportion of "green" revenue multiplied by the current number of employees.

$$Y_{cell} = \sum_{1}^{c} w_i y_i$$

Weighting Class Adjustment for Nonresponse – To mitigate possible bias arising from nonresponse, weighting class adjustments to the weights will be made. Initial plans are to define State x industry sector cells as the weighting classes. With modified PPES sampling, an adjustment based on employment will be used. Let e_i be an establishment's employment size used for allocation and sample selection (not the current size). Of c sample units in a weighting class, suppose that r respond. Weights of respondents are increased to "cover" the missing data of nonrespondents. Additionally, modifications are made to account for establishments that are out-of-scope (n_{oos}) or out-of-business (n_{oob}).

$$E_{class} = \sum_{1}^{c-noos-noob} w_i e_i$$

$$E'_{class} = \sum_{1}^{r} w_i e_i \qquad w'_i = w_i * (E_{class} / E'_{class})$$

Benchmarking – Benchmarking to in-scope (ANAICS) employment for national ANAICS and State x industry sector is planned. Some protection would be provided against coverage shortcomings. Since "green" economic activity is anticipated to be a small part of most industries, little variance-reduction benefit would be expected from benchmarking. The classes may be structured differently, but a weight adjustment similar to weighting class adjustment for nonresponse can be computed. Instead of employment e_i at the time of sampling, use the latest employment E_i; that is control the sample weighted estimates to known population values obtained from an updated QCEW. A ratio adjustment is computed from the unweighted "known" employment total for the class (the sum of E_i over the population N) and the sample estimate that can be made from the respondents weighted by w_i'.

$$BME_{class} = \sum_{1}^{N} E_i \qquad BME'_{class} = \sum_{1}^{r} w'_i E_i \qquad w''_i = w'_i * (BME_{class} / BME'_{class})$$

For totals, simple weighted estimates can be made using the weights w_i" and the responding establishment values y_i for a characteristic. The sum in the formula is restricted to the r respondents that contribute to an estimate y of an unknown population characteristic Y.

$$y \dot{\iota} \sum_{1}^{r} w_i'' \, y_i$$

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

2c. Reliability

The estimation of sample variances will use a replication methodology similar to that used by Current Employment Statistics (CES) and the Job Openings and Labor Turnover Survey (JOLTS). Balanced Half Sampling (BHS) uses half samples of the original sample and calculates estimates using those half samples. Balanced Repeated Replication (BRR) modifies the technique by using the entire sample for making replicate estimates but by perturbing the weights of half samples in a systematic fashion using a Hadamard matrix. The sample variance is calculated by measuring the variability of the estimates made from these replicates. (For a detailed mathematical presentation of this method as applied to estimates from the Current Employment Statistics Survey, see Handbook of Methods, Chapter 2, pages 16, Bureau of Labor Statistics, updated 12/2010, or http://www.bls.gov/opub/hom/pdf/homch2.pdf.) A method with a different Hadamard matrix is planned for GGS.

The standard weight perturbation uses a factor of ½ to either increase or decrease the weight for an establishment when making an estimate for the αth replicate.

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

where

$\gamma = 0.5$ -- perturbation factor;

$G_i = \pm 1$ -- random groups indicator;

$h^{(\alpha)}$ -- element of the Hadamard matrix (αth row for a given column);

f -- sample fraction;

w_i -- selection weight

As mentioned beginning of Part B, this is a new survey for BLS, thus at present estimates of standard errors are not available. These estimates will be provided after the data from the first sample is tabulated.

For each replicate α, an estimate can be made, for example of a total Y. If there are A replicates, each replicate can be compared to an estimate Y' made from the entire sample (using original weights) and the following formula used to calculate an estimated variance.

$$var_{BRR}(Y') = \frac{1}{A\gamma^2} \sum_{\alpha=1}^{A} \left(Y'^{(\alpha)} - Y'\right)^2$$

2e. Specialized Procedures
Extensive research is being conducted on GGS sampling methodology and coordination with Occupational Employment Statistics. As much as possible, the GGS sample will be drawn as a subsample of OES, but in some States/industries that will require augmenting the existing OES sample.

2f. Data Collection Cycles
GGS data will be collected annually for about 120,000 establishments. The sample will be divided into 3 panels of about 40,000 establishments each. After start up, one panel will be dropped and replaced by another panel each year. A new panel will be included for data collection each year. Following is a schematic of the planned panel rotation. Start-up panels are labeled panel1, panel2, and panel3. Replacement panels are labeled panel4, panel5, etc.

Year   Panels Included

2011   panel1 panel2 panel3

2012   panel2 panel3 panel4

2013   panel3 panel4 panel5

2014   panel4 panel5 panel6

2015   panel5 panel6 panel7

3. Methods to Maximize Response Rates and Non Response Adjustment

3a. Maximize Response Rates
Before mailing, standard BLS address refinement procedures are implemented. Then, employers are mailed an advanced letter, followed by the cover letter and data collection form. The cover letter pledges confidentiality and explains the importance of the survey and the need for voluntary cooperation. There are two follow-up mailings. Depending on resources, interviewers will start calling some establishments who have not responded after the second mailing and others after a third mailing and attempt to enroll them into the survey. The follow-up contact is especially important for unique certainty establishments in a State or industry. Non-respondents and establishments that are reluctant to participate are re-contacted by an interviewer especially trained in refusal aversion and conversion.

The response rate is unknown, but the data collection schedule will be somewhat similar to that of the Occupational Employment Statistics survey. The OES establishment response rate 2009 to the current time averaged 78.2% and the response rate based on employment averaged 74.5%. The general formulas are given here for weighted response weights that take into account the original base weights

w_i (inverses of sampling probabilities). Here r is the number of responses and n' is the original sample size n reduced by the number of out-of-business establishments. Out-of-business establishments are nearly 100% identified by survey procedures but are not counted as responses in the formulas. The employment size used for establishment i during sample selection is e_i.

$$Weighted\ Establishment\ Response\ Rate = \sum_1^r w_i / \sum_1^{n'} w_i$$

$$Weighted\ Employment\ Response\ Rate = \sum_1^r w_i e_i / \sum_1^{n'} w_i e_i$$

3b. Response and Non-Response Analysis Surveys
BLS is also requesting OMB clearance to conduct a series of response and non-response analysis surveys to assess the quality of the data in terms of response and non-response errors and biases. The total maximum sample size for all follow-up surveys is 3,000 establishments each of 20 minutes or a total of 1,000 hours. The data collection and analysis for these surveys is contingent upon available resources.

3c. Non Response Adjustment
Prior to calculating estimated response rates, preliminary editing procedures will flag questionable data. The extent and nature of item nonresponse is not known. It is likely that responses that are incomplete after follow-up contact will be dropped.

Weighting class adjustments to sampling weights w_i will be made to partially compensate for the bias that arises if nonresponse is ignored (see Estimation Procedure and Maximize Response Rates sections). In the simplest form, the entire sample is divided into weighting classes based on strata or a simple subdivision of the population. The default for GGS will be weighting classes defined State x industry sector, and analysis of responses will determine if more complexity is needed.

As mentioned above, response and non-response analysis surveys are planned to assess both response and non-response errors and biases, if any. The first one is planned two months after the data collection begins.

4. Tests
Prior to fielding the first of four data collection tests, BLS conducted feasibility interviews and cognitive testing on the potential survey questions, and the availability and ease of collecting GGS data from respondents. These tests guided the design of the data collection forms. Once it was determined the data were collectable, the GGS survey conducted four different data collection tests over the spring and summer of 2010. After each data collection test, 50 telephone follow-up phone calls were made to respondents who completed the forms, and amongst those 50 telephone calls, approximately half of the calls were made to respondents who correctly completed the form, and the remaining calls were made to people who completed the form incorrectly. Additionally, during non-response telephone calls as part of the data collection efforts, telephone interviewers would note comments made by respondents about the data collection form.

The data collection forms being submitted for approval have been field tested in the final data collection test and received the highest rating of satisfaction of the forms being tested. The GGS staff also consulted with cognitive researchers within BLS to ensure these forms are easy to understand and complete.

5. Statistical and Analytical Responsibility

The GGS survey is in the BLS Office of Employment and Unemployment Statistics. Ms. Shail Butani, Chief, Statistical Methods Staff is responsible for the statistical aspects of the GGS survey and can be reached on 202-691-6347. Mr. Richard Clayton, Chief, Division of Administrative Statistics and Labor Turnover, has overall program oversight.

6. References

Bureau of Labor Statistics' Handbook of Methods, Chapter 2, pages 16, Bureau of Labor Statistics, updated 12/2010. (http://www.bls.gov/opub/hom/pdf/homch2.pdf)

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

Attachments for Section B

1. Frame Summary by Major Industry Sector and Ownership
2. Frame Summary by State and Ownership
3. Sample Count Summary by Major Industry Sector and Ownership
4. Sample Count Summary by State and Ownership