Supporting Statement for Request for OMB Approval
Occupational Employment Statistics Green Technologies and Practices
Data Collection Clearance

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

The Green Technologies and Practices (GTP) survey is an occasional survey conducted by the BLS.

1(a) Respondent Universe

The universe for this survey includes private and government establishments in the 50 States and the District of Columbia. The sampling frame is the Quarterly Census of Employment and Wages (QCEW) Longitudinal Database (LDB) maintained by the Bureau of Labor Statistics. The QCEW program produces a comprehensive tabulation of employment and wage information for workers covered by State unemployment insurance (UI) laws and Federal workers covered by the Unemployment Compensation for Federal Employees (UCFE) program. Included are data on the number of establishments, monthly employment, and quarterly wages, by industry based on the North American Industrial Classification System (NAICS), by county, and by ownership sector, for the entire United States. This information is provided for over 9 million business establishments of which about 6.9 million are in the scope of this survey. (Out of scope establishments include private households, establishments with zero employment, and establishments in outlying areas.)

1(b) Sample

Scope–The GTP will measure occupational employment and wages of workers engaged in green practices in establishments in the 50 States and the District of Columbia. The survey covers all NAICS industries except NAICS 814, private households.

Sample Size–The sample size is approximately 35,000 establishments. The following table shows the estimated number of universe units, sampled units, and responding units (assuming a 75 percent response rate, see 3(a)) for all in-scope NAICS for the GTP survey:

Table 1: Universe and Sample Size Summary

Survey	NAICS Coverage	Responding Units	Sample Units	Universe Units
FY 2012	All NAICS Except 814	26,251	35,000	6,900,000

Stratification–Units on the sampling frame will be stratified by Census Region and by two-digit NAICS industry sector.

2(a) Sample Design

The GTP will have a probability-based sample aimed at satisfying data needs at both the national and Census Region/2-digit industry level. The basic sampling unit is an establishment. The most recent data from the 2011 GTP survey show that about 70 percent of responding establishments from the QCEW frame reported green technologies or practices at their business. 11.6 percent of establishments reported employees meeting our green definition. BLS used a rare-subpopulation sampling method to include a larger number of green establishments. To accomplish this, BLS has conducted research on establishments known to be green and has compiled its own green list through web research and the use of other known green business organization lists. This list contains about 32,000 establishments and will be incorporated into the sample in separate strata and sampled with a higher probability than other establishments.

Allocation method–A power allocation based on the square root of stratum size, defined by frame employment within each stratum will be used to allocate the sample to each Census Region/2-digit NAICS stratum. This is a compromise between 1) allocating an equal sample size to each stratum which is optimal where estimates for individual strata are of prime importance and 2) a straight proportional allocation based on stratum size which would be optimal for estimation of high-level data.

Sample Selection–Within each stratum, the sample is selected using modified probability proportional to estimated employment size (PPES). The estimated employment size for an establishment will be the maximum employment over the last 12 months of data available from the LDB. Units with average employment of zero over the last 12 months are excluded. Each sampled establishment is assigned a sampling weight equal to the reciprocal of its probability of selection in the sample.

2(b) Estimation Procedures

GTP estimators of total employment 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 number of employees who spend more than 50% of their time engaged in green practices.

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

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 Census Region by industry sector cells as the weighting classes. With 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}).

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

$$E'_{class} = \sum_{1}^{r} w_i e_i$$

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

$$w'_i = w_i \left(\frac{E_{class}}{E'_{class}} \right)$$

$$w'_i = w_i * (E_{class} / E'_{class})$$

Benchmarking–Benchmarking to in-scope employment for national and Census Region by 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, shown as BME_{class} below) and the sample estimate that can be made from the employment values reported by the respondents to the survey [E_i^r weighted by w'_i (sum from 1 to R below), shown as BME'_{class} below].

$$w''_i = w'_i * (BME_{class} / BME'_{class})$$

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

$$w''_i = (BME_{class} / BME'_{class})$$

For totals, simple weighted estimates can be made using the responding establishment values y_i for a characteristic and the final weights (fw_i below), which are the product of the sampling weights, the nonresponse adjustment factors, and the benchmark weights from above. The sum in the second formula below is restricted to the R respondents that contribute to an estimate y of an unknown population characteristic Y.

$$fw_i = w_i * w'_i * w''_i$$

$$y = \sum_{1}^{R} fw_i y_i$$

Wage Estimation–Mean wage and median wage estimates will be calculated for occupations within Census Region/two-digit industry cells. Wage rate data will be collected in broad wage bands instead of exact data points. To approximate median wage rates, a uniform distribution

within wage intervals is assumed and a simple linear interpolation between the endpoints of the wage intervals is used.

Because wage rate data are collected in broad wage bands instead of exact data points, the standard error for each mean wage rate estimate is calculated using a components model. This model accounts for the variability of both the observed and unobserved components of wage rate data. A traditional ratio variance estimator will be used to account for the variability observed in the collected GTP wage data.

2(c) 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, BLS Chapter 2, pages 8-9, Bureau of Labor Statistics, 200 or http://www.bls.gov/opub/hom/homch2.htm.) A method with a different Hadamard matrix is planned for GTP.

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 α^{th} replicate.

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

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);

w_i -- selection weight

For each replicate α, an estimate $Y'^{(\alpha)}$ 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.

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

Below are resultant sample sizes and CVs by 2-digit NAICS from the first GTP survey.

Industry			
NAICS2	Sector Name	Previous GTP National Sample Size	Previous GTP National CV
11	Agriculture/Forestry/Fishing/Hunting	730	33.60%
21	Mining	390	39.57%
22	Utilities	585	25.60%
23	Construction	1,517	18.45%
31	Manufacturing, Non Durable	888	11.03%
32	Manufacturing, Non Durable	1,090	11.03%
33	Manufacturing, Durable	1,799	11.03%
42	Wholesale Trade	1,406	22.65
44	Retail Trade	1,729	24.85%
45	Retail Trade	1,220	24.85%
48	Transportation & Warehousing	1,097	18.94%
49	Transportation & Warehousing	820	18.94%
51	Information	1,001	39.53%
52	Finance & Insurance	1,665	41.64%
53	Real Estate & Rental/Leasing	912	31.94%
54	Prof/Scientific/Tech. Services	1,947	21.97%
55	Management of Co & Enterprises	844	35.43%
56	Admin/Waste/Remediation Services	1,939	20.55%
61	Education Services	3,007	37.30%
62	Health Care/Social Assistance	3,911	21.16%
71	Arts/Entertainment/Recreation	1,043	15.53%
72	Accommodation/Food Services	7,878	25.30%
81	Other Services	1,146	23.76%
92	Public Administration	2,303	18.88%
All	Across All Industries	34,867	6.86%

2(d) Data Collection Cycle

The Green Technologies and Practices Survey is an occasional data collection effort.

3(a) Maximizing Response

A goal of the GTP survey is to achieve a 75 percent response rate. The 2011 GTP survey obtained an adjusted usable response rate of 70 percent.

The survey protocol includes the initial mailing of the survey form, return envelope, cover letter, and web reporting instructions. A non-response prompt postcard will be mailed to respondents within four weeks of the initial mailing. Approximately three weeks following the postcard

mailing a third survey mailing including the survey form, return envelope, and web reporting instructions will be sent. Telephone data collection for all non-respondents will begin approximately two weeks following the third and final mail out and continue until the end of data collection.

The GTP is a voluntary survey. Every effort will be made to maximize response rates to achieve the 75-percent goal by:

- Utilizing a web-based collection instrument. Research determined that web-based reporting is important to the success of this survey due to the subject matter.
- Researching and obtaining e-mail addresses and contact names for sampled units where possible. When contact names are not available, our research suggests the following guidelines be used in addressing the forms: "Proprietor/Owner" for firms with 10 or fewer employees; "General Manager" for larger firms with less than 100 employees; and "Human Resources Manager" for firms with more than 100 employees.
- Conducting extensive address refinement to ensure that the survey solicitation package reaches the correct establishment in a timely manner.
- Providing each sampled unit with a cover letter explaining the importance of the survey and the need for voluntary cooperation.
- Giving each private sector sample unit the Bureau's pledge of confidentiality.
- Sending each nonresponding unit two additional mailings after the initial mail-out (and contacting nonresponding units by telephone).
- Using status reports and control files to identify cells with low response rates and targeting nonresponse follow-up activities and telephone data collection for key respondents in these cells
- Stressing to respondents that assistance is available to help them complete the survey form.
- Using a respondent web page that provides detailed information about responding to the GTP survey, including contact information for those needing assistance and a link to on-line data reporting.
- Increasing the use of electronic and telephone collection in order to allow the respondent to provide information in a way that is most convenient to them.

3(b) Nonresponse Adjustment

Prior to calculating estimates, preliminary editing procedures will flag questionable data. The extent and nature of item non-response is not known. Responses that are incomplete after follow-up contact will either be made unusable (treated like unit non-response) or imputed. In the 2011 GTP survey 1,649 units (approximately 7%) out of 22,484 usable responses were imputed.

Weighting class adjustments to sampling weights w_i will be made to partially compensate for the bias that arises if non-response is ignored (see Estimation Procedure section). 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 GTP will be weighting classes defined by Census Region x industry sector, and analysis of responses will determine if more complexity is needed.

Given a hypothesized response rate (75%) that is less than optimal (80 – 100%), nonresponse bias in key estimates may be assessed through multivariate models that incorporate data from, primarily, two sources. The first source of data is the sample frame data. The sampling frame for the GTP survey, the LDB, contains establishment size and industry information, as well as age of the sampled establishment. The second source, sometimes referred to as "paradata," describes important aspects of the administration of the survey. These data include variables that may describe some of the following: the number and type of contact attempts and interview forms or reminders mailed to each establishment, the time in the field, the job title of the respondent within the sampled establishment, the number of different persons in the establishment that have been contacted, whether the form was sent to a central office, any expressions by the contacts within the establishment indicating unwillingness or refusal to complete the survey, and any interviewer comments recorded for the record. Nonresponse propensity using data from these two sources will be assessed, possibly through multivariate logistic regression models, within each mode of survey administration (telephone, internet, and mail) and across the modes of survey administration. The propensity to respond will then be examined in terms of its contribution to bias on key survey variables, such as the number and type of green activities and the number and type of green jobs.

3(c) Confidentiality

Before estimates are released to the public, they must first be 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 private sector sample unit that the BLS will not release its employment data to the public in a manner that would allow others to identify the unit. If a green estimate fails confidentiality screening, the estimate is suppressed.

4. Developmental Tests

A Response Analysis Survey (RAS) was conducted at the conclusion of the 2011 GTP Survey under OMB Clearance 1220-0184. A series of questions related to the respondents' interpretation of the survey's questions, their ability to provide accurate information, and the amount of time it took to complete the questionnaire, either on paper or on-line, were asked. Eighty-one percent of web respondents reported that the survey was either easy or very easy to complete, while less than 2 percent reported that the survey was difficult or very difficult. Seventy-three percent of mail respondents reported that the survey was either easy or very easy to complete, while less that 6 percent reported that the survey was difficult or very difficult. The median response time reported for completing the survey on-line was 20 minutes, while the median response time for completing the mail survey was 30 minutes.

5. Statistical and Analytical Responsibility

Ms. Shail Butani, Chief, Statistical Methods Division of the Office of Employment and Unemployment Statistics, and Mr. George Stamas, Chief, Division of Occupational Employment

Statistics, Office of Employment and Unemployment Statistics are responsible for the statistical and analytical aspects of the GTP program respectively. Ms. Butani can be reached on 202-691-6347. Mr. Stamas can be reached on 202- 691-6350. Additionally, BLS seeks consultation with other outside experts on an as needed basis.

6. References

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

Bureau of Labor Statistics' Handbook of Methods, Chapter 3, Bureau of Labor Statistics, 2008 (http://www.bls.gov/opub/hom/homch3_a.htm)

Bureau of Labor Statistics' Handbook of Methods, Chapter 2, pages 8-9, Bureau of Labor Statistics, 2004 (http://www.bls.gov/opub/hom/homch2.htm.)

Kalton, Graham and Anderson, Dallas W. (1986). Sampling Rare Populations. *Journal of the Royal Statistical Society, A.* 149:69-82.

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

Lohr, Sharon L. and Rao, J. N. K. (2006), Estimation in Multiple-Frame Surveys, *Journal of the American Statistical Association*. 475:1019-1030.

OES State Operations Manual, Bureau of Labor Statistics, Internal Document, 2008.

Piccone, David and Stetser, Marie. (2009) *National Sample Reallocation for the Occupational Employment Statistics Survey,* 2009 Joint Statistical Meetings.

Technical Notes for May 2011 OES Estimates, Bureau of Labor Statistics, 2011 (http://www.bls.gov/oes/current/oes_tec.htm).

Technical Notes for Aug 2011 GTP Estimates, Bureau of Labor Statistics, 2011 (http://www.bls.gov/gtp/survey_methods.htm).