Survey Methods and Reliability Statement for the August 2011 Green Technologies and Practices Survey

I. Introduction

The Green Technologies and Practices (GTP) survey is a special survey of business establishments designed to measure the use of technologies and practices that lessen the environmental impact of an establishment's production processes. The survey also collects occupational employment and wage data for wage and salary workers who spent more than half of their time involved in green technologies and practices during the survey reference period, the pay period including August 12, 2011.

The GTP survey collects information on the BLS process approach to measuring green jobs: jobs in which workers' duties involve making their establishment's production processes more environmentally friendly or use fewer natural resources. More information about the BLS green jobs initiative is available from the green jobs homepage at www.bls.gov/green.

The GTP survey draws its sample primarily from the Quarterly Census of Employment and Wages (QCEW) state unemployment insurance (UI) files. About 6.7 million establishments in the 50 states and the District of Columbia were stratified by Census region and 2007 North American Industry Classification System (NAICS) 2-digit industry sector. From this sampling frame, a probability sample of about 35,000 establishments was selected.

Survey forms were mailed to sampled business establishments. About 70 percent of sampled establishments responded to the GTP survey. Forty-seven percent of respondents provided data by telephone, 37 percent returned the survey form by mail, and the remainder responded by email, fax, or on the internet.

Respondents were asked whether or not they used each of six green technologies and practices during the pay period that included August 12, 2011. They were also asked to provide the number of employees who spent more than half of their time involved in green technologies and practices during the reference period. For such workers, respondents were asked to provide job titles and brief job descriptions, as well as the number of workers, by occupation, in each of 12 specific wage intervals. The wage intervals were defined in terms of both hourly rates and the corresponding annual rates, where the annual rate for an occupation is calculated by multiplying the hourly wage rate by a typical work year of 2,080 hours. Respondents were instructed to report part-time workers at their hourly rates. Full-time workers could be reported by either hourly rates or annual salaries, depending on how the worker was paid.

II. Definitions

Green technologies and practices are technologies and practices that lessen the environmental impact of an establishment's production processes. Employers were asked whether they had used each of the six green technologies and practices listed below during the reference period. Examples were provided of the types of technologies and practices included in each of the six categories.

Energy from renewable sources and energy efficiency

1. Generate electricity, heat, or fuel from renewable sources primarily for use within the establishment.

Examples of renewable sources:

- Wind
- Geothermal
- Ocean
- Landfill gas
- Biomass
- Solar
- Hydropower
- Municipal solid waste

2. Use technologies or practices to improve energy efficiency within the establishment.

Examples:

- Energy Star rated appliances
- Occupying a LEED (Leadership in Energy and Environmental Design) certified building
- Energy efficient lighting
- Programmable thermostats
- Cogeneration (combined heat and power)
- Energy efficient manufacturing equipment

Greenhouse gas reduction and pollution reduction and removal

3. Use technologies or practices in operations to reduce greenhouse gas emissions through methods other than renewable energy generation and energy efficiency.

Examples:

- Purchase and use of carbon offsets
- Promotion and/or subsidy of alternative forms of transportation for employees, such as carpools, fuel efficient vehicles, cycling, or mass transit
- Implementation of a telework program for employees

4. Use technologies or practices to either reduce the creation or release of pollutants or toxic compounds as a result of operations, or to remove pollutants or hazardous waste from the environment.

Examples of pollutants or toxic compounds:

- Carbon monoxide
- Sulfur dioxide
- Chlorofluorocarbons (CFCs)
- Nitrogen oxides
- Chlorinated hydrocarbons
- Herbicides or pesticides
- Heavy metals
- Radioactive contamination

Recycling and reuse and natural resource conservation

5. Use technologies or practices to reduce or eliminate the creation of waste materials as a result of operations.

Examples:

- Collecting and reusing or recycling waste
- Managing wastewater
- Composting solid waste
- Remanufacturing

6. Use technologies or practices in operations to conserve natural resources, excluding the use of recycled inputs in production processes.

Examples:

- Managing land resources
- Managing storm water
- Conserving soil, water, or wildlife
- Implementing organic agriculture or sustainable forestry practices

An ***establishment*** is generally a single physical location at which economic activity occurs (e.g., store, factory, restaurant, etc.). Each establishment is assigned a 6-digit NAICS code. When a single physical location encompasses two or more distinct economic activities, it is treated as two or more separate establishments if separate payroll records are available and certain other criteria are met.

Employment is defined as the number of full- and part-time workers who are paid a wage or salary, including paid owners, officers, and staff of incorporated firms and workers temporarily assigned to

other locations. The survey does not include the self-employed or owners, partners, and proprietors of unincorporated firms; unpaid family workers; workers on unpaid leave; workers not covered by unemployment insurance; and contractors and temporary agency employees not on the sampled establishment's payroll.

GTP employment refers to the number of jobs in which workers spend more than half of their time involved in green technologies and practices. Employees were considered to be involved in green technologies and practices if they were researching, developing, maintaining, using, or installing green technologies and practices, or training the establishment's workers in these technologies and practices.

An ***occupation*** is a set of activities or tasks that employees are paid to perform. Workers are classified into occupations based on their job duties and, in some cases, on the skills, education, and/or training required. Workers with similar job duties are classified in the same occupation, regardless of the industry in which they are employed. The GTP survey uses the 2010 Standard Occupational Classification (SOC) system to classify workers into occupations.

Wages are money that is paid or received for work or services performed in a specified period. Wages for the GTP survey are straight-time, gross pay, exclusive of premium pay.

Included in the collection of wage data are:

- Base rates
- Commissions
- Cost-of-living allowances
- Deadheading pay
- Guaranteed pay
- Hazard pay
- Incentive pay
- Longevity pay
- Over-the-road pay (mileage)
- Piece rates
- Portal-to-portal rates
- Production bonuses
- Tips

Excluded from the wage data are:

- Attendance bonuses
- Back pay
- Clothing allowances
- Discount
- Draw
- Holiday bonuses

- Holiday premium pay
- Jury duty pay
- Meal and lodging payments
- Merchandise discounts
- Nonproduction bonuses
- On-call pay
- Overtime pay
- Perquisites
- Profit-sharing payments
- Relocation allowances
- Severance pay
- Shift differentials
- Stock bonuses
- Tool/equipment allowances
- Tuition repayment
- Uniform allowance
- Vacation pay
- Weekend premium pay
- Year-end bonuses

III. Sample design

Frame Creation

The GTP sampling frame has about 6.7 million in-scope establishments, which includes private and government establishments in the 50 states and the District of Columbia. The frame is developed primarily from the state Quarterly Census of Employment and Wages (QCEW) files for the 3rd quarter of 2010. The QCEW includes all business establishments subject to unemployment insurance (UI) tax. In addition to the QCEW data file, a railroad sampling frame is used.

BLS also conducted research to identify 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 31,000 establishments and was used as a separate frame in order to target these establishments; this will be referred to as the green frame in later sections.

Stratification

Establishments on the frame are stratified by Census region and 2-digit industry sector (NAICS).

- Geography—There are four Census regions: Northeast, Midwest, South, and West.

- Industry—There are 20 2-digit industry sectors:

11	Agriculture, forestry, fishing, and hunting
21	Mining
22	Utilities
23	Construction
31-33	Manufacturing
42	Wholesale trade
44-45	Retail trade
48-49	Transportation and warehousing
51	Information
52	Finance and insurance
53	Real estate and rental and leasing
54	Professional, scientific, and technical services
55	Management of companies and enterprises
56	Administrative and support and waste management and remediation services
61	Educational services
62	Health care and social assistance
71	Arts, entertainment, and recreation
72	Accommodation and food services
81	Other services, except public administration [private households (814) are excluded]
92	Public administration

Sample Allocation

The sample of approximately 35,000 units is allocated into the strata defined above. About 33,000 establishments are allocated to the GTP sample from the QCEW frame. About 2,000 establishments are allocated from the green frame mentioned above. The sample is allocated according to the following formula:

$$n_h = N * \frac{\sqrt{X_h}}{\sum_{h=1}^{H} \sqrt{X_h}}$$

h = a sampling cell defined by Census Region, 2-digit NAICS industry (*h* = 1, 2, …, *H*)

N = the national sample size

X_h = the total frame employment of establishments in sampling cell *h*

n_h = the number of sample units allocated to sampling cell *h*

For the cell employment X_h, the employment for each establishment is defined as the maximum of that establishment's 12 monthly employment values. The maximum employment is used to define the size of an establishment in order to eliminate the need for any adjustments due to seasonality. Units with employment less than or equal to 10 are treated as if they have 10 employees for this procedure.

Selection

Within each stratum, the sample is selected using a modified probability proportional to estimated employment size (PPES) method. The employment size for an establishment is determined by the maximum employment over the prior 12 months of data available from the QCEW. Units with employment of zero over the 12 months were excluded.

Assignment of Sample Weights

Each sampled establishment is assigned a sampling weight equal to the reciprocal of its probability of selection in the sample. These weights are later adjusted when nonresponse and benchmark employment factors are taken into account. These weights are computed so that the sample will represent the entire universe of establishments.

IV. Estimation

Horvitz-Thompson Estimator

A Horvitz-Thompson (HT) estimator is used to estimate GTP employment and the number of establishments reporting green technologies and practices. Establishments' reported total employment and GTP employment figures are used to calculate the employment estimates. Every establishment has a final weight that is a combination of sampling weights, benchmark factors, and nonresponse adjustment factors. In order to calculate estimates for the whole population, these final weights are multiplied by the corresponding variables, such as reported employment, GTP employment, or green technologies and practices reported.

The estimation levels for the GTP survey are:

- National
- Census region
- 2-digit NAICS sector
- Occupations, detailed and groups

For estimation cell *h*, estimates are computed using the following formula:

$$\hat{y}_h = \sum_{i \in h} (fw_i * x_i)$$

$\boldsymbol{fw_i}$ = final weight for establishment *i*

$\boldsymbol{x_i}$ = reported value for establishment *i*

This formula is used to compute many different kinds of estimates, such as green employment and number of establishments reporting green technologies and practices.

Nonresponse Adjustment

For a variety of reasons, some sampled establishments either fail to respond or fail to provide complete, usable information on the survey form. Both types of establishments are considered nonrespondents, but are handled differently. Establishments that do not report occupational wage data and green practices data are considered unit nonrespondents. Establishments that have missing or zero reported employment and do not have employment on the QCEW frame are also considered unit nonrespondents. The nonresponse adjustment factors are calculated to account for these units that did not provide usable response information. On the other hand, establishments that report occupations and their totals but fail to report corresponding occupational wage data or establishments that report only part of the green practices data are called partial nonrespondents, and are imputed according to procedures detailed in a later section.

Unit nonresponse adjustment was conducted at the Census region/2-digit NAICS/Size class level. If there are not enough sample units in the cell, then size classes are collapsed until we get a sufficient number of units in the cell for nonresponse adjustment calculations. The nonresponse factors are calculated using the following formula:

$$NRAF_{i,h} = \frac{\sum_{i \epsilon S_h} w_i * Max\_EMP_i}{\sum_{i \in R_h} w_i * Max\_EMP_i}$$

Max_EMP_i = maximum QCEW employment over 12 months of unit *i* in nonresponse adjustment cell *h*

w_i = sampling weight of unit *i* in nonresponse adjustment cell *h*

S_h = sampled establishments in cell *h*

R_h = usable respondents in cell *h*

Benchmarking

The benchmark process ensures that the employment estimates are consistent with employment figures from the QCEW program. Benchmarking is performed at the Census region/2-digit NAICS level and is done after nonresponse adjustment. The weights used in the benchmark calculation are modified by nonresponse adjustment factors (NRAF). The auxiliary variable for the estimator is the August 2011 employment from the QCEW.

The benchmark factors are calculated using the following formula:

$$BMF_{i,h} = \frac{\sum_{i \epsilon N_h} Bmk\_EMP_i}{\sum_{i \in R_h} w_i \ * \ NRAF_i \ * Rpt\_EMP_i}$$

Bmk_EMP_i = August QCEW employment for unit *i* in cell *h*

Rpt_EMP_i = reported employment for unit *i*

w_i = sampling weight of unit *i*

$NRAF_i$ = nonresponse adjustment factor of unit *i*

$BMF_{i,h}$ = benchmark factor of unit *i* in adjustment cell *h*

N_h = frame establishments in cell *h* (2011 3rd quarter QCEW)

R_h = usable respondents in cell *h*

Imputation

Two different types of imputation methods are used for missing data. Nearest neighbor imputation is used for green technologies and practices and wage distribution imputation is used for reported occupations that are missing their occupational wage distributions.

The ***nearest neighbor imputation*** method is used to fill in missing green technologies and practices data for partial nonrespondents. A donor pool of respondents is found that most closely resembles each nonrespondent by geography, industry, and size. The search begins at the Census region/5-digit NAICS/size class level; if a suitable donor pool is not found, the search continues in a hierarchical manner by expanding the geography, industry, and size class. In addition to the hierarchy, two additional methods are used to help conceive a donor pool: the g*reen activity match* and the e*mployment distance function* methods. The green activity match is used when a unit reports answers to only some of the green technologies and practices questions. The employment distance function is used when a unit reports no answers to all of the green technologies and practices questions. In this case, a distance is computed using the reported employment and reported green employment to find the

closest donors. Once the donor pool is created, the partial nonrespondent is imputed using the donors' combined answers to the green technologies and practices questions.

The ***wage distribution imputation*** method is used to impute wage distributions for units that report occupations and employment but not the wage distributions. A donor pool is selected in a similar way as described above in the nearest neighbor imputation. The search for a donor pool here starts with establishments reporting wage information for the occupation in the same MSA/4-digit NAICS/size class level. If a suitable donor pool is not found, the search continues in a hierarchical manner by expanding geography, industry, and size class. Once a donor pool is created, the distribution across wage intervals is computed using the weighted occupational employment of respondents that report that occupation. The distribution is then used to prorate the nonrespondent's occupational employment total across the wage intervals.

Mean and Median Wage Estimates

Since the GTP survey collects wage data by wage intervals rather than by wage rate, special procedures are needed to produce mean and median wage estimates.

Mean Wage Estimates

Mean wage estimates are calculated using a weighted mean of the 12 wage intervals (A through L). In order to estimate this, means for the individual wage intervals are needed. These are calculated using harmonic means for 11 of the 12 wage intervals. The interval mean for the highest, open-ended interval is calculated based on data from the BLS National Compensation Survey. For the lowest wage interval, state-specific harmonic means are calculated that incorporate each state's minimum wage.

The harmonic mean used to compute each interval mean is

$$\overline{x_j} = \frac{2 * x_1 * x_2}{x_1 + x_2}$$

x_1 = the lower bound of interval *j*
x_2 = the upper bound of interval *j*

The mean wage rate for occupation O in estimation cell Ω is:

$$\bar{x}_{\Omega,o} = \frac{\sum_{i=1}^{n_{\Omega,o}} \sum_{j=A}^{L} \left(\overline{x_j} * wage\_range\_emp_j * Final\_wgt_i \right)}{\sum_{i=1}^{n_{\Omega,o}} \sum_{j=A}^{L} \left(wage\_range\_emp_j * Final\_wgt_i \right)}$$

$\overline{x_j}$ = mean wage rate for interval *j* (*j*= A, B,…,L)

$wage\_range\_emp_j$ = reported occupational employment under wage interval *j*

$Final\_wgt_i$ = final weight of unit *i*

$n_{\Omega,o}$ = number of units with occupation O in estimation cell Ω

Median Wage Estimates

The median or 50th percentile hourly wage rate for an occupation is the wage where 50 percent of all workers earn that amount or less and where 50 percent of all workers earn that amount or more. The wage interval containing the median hourly wage rate is located using a cumulative frequency count of estimated employment across all wage intervals. After the targeted wage interval is identified, the median wage rate is then estimated using a linear interpolation procedure. In the GTP survey, weighted median estimates are calculated where establishment weights are taken into consideration.

Variance

The Green Technologies and Practices survey uses Fay's modified Balanced Repeated Replication method to calculate variances. This method splits the sample into halves and multiplies each unit's sampling weight by 0.5 or 1.5 based on values from a Hadamard matrix. All estimates are then computed with these modified sampling weights. This is repeated *γ* times, with each replicate producing a different half-sample and a different set of estimates. These replicates are used to estimate sample variances using the formula below:

$$\hat{V}(\hat{z}) = \frac{1}{\gamma*(1-K)^2} \sum_{i=1}^{\gamma} (\hat{z}_i^{\alpha} - \hat{z})^2$$

$\hat{z}_i^{\alpha}$ = i^{th} replicate estimate of population parameter *Z*

$\hat{z}$ = sample-based estimate of population parameter *Z*

$K = 0.5$

γ = number of replicates (140 in the GTP survey)

V. Quality Control

The GTP survey underwent rigorous design and response testing prior to production. Cognitive interviews were conducted with establishments thought to have green technologies and practices to further BLS's understanding of environmental terminology and relevance. A feasibility study was conducted to assess both the understanding of the survey's language and firms' ability to provide the requested data. Five test panels were conducted to refine the survey procedures and collection instruments: mail survey form, fax survey form, email survey form, and internet collection form. Response analysis surveys were conducted on a small number of respondents and nonrespondents in each of the five test panels to further understand respondents' and nonrespondents' reactions to the survey questions and their reasons for response or nonresponse.