B. Collection of Information Employing Statistical Methods

1. The Producer Price Index is an on-going survey of net transaction prices received by producers. For a general description of the methodology for the PPI, see the BLS Handbook of Methods chapter 14. The current universe for the PPI survey consists of roughly 5.0 million establishments comprising the covered portions of the mining, manufacturing, forestry, utility, and non-goods producing sectors. About 6,400 respondents are undergoing initiation in any given year, while 26,250 respondents provide monthly data pertaining to 105,000 price quotations. The replenishment and rotation of respondents within the PPI occurs at a rate of approximately one-seventh each year. The list containing the universe of all producing establishments comes mostly from Unemployment Insurance (UI) files provided by state agencies. Supporting information and alternative frames may be obtained from other sources, if they are deemed to be more accurate. Total initial cooperation is about 80 percent. Roughly 20 percent of establishments slated for initiation into the PPI refuse to cooperate. Frame error further reduces the PPI's initial collection rate, while respondent attrition over the life of the sample affects the long-term cooperation rate.

2. The PPI survey is based on a stratified probability-proportional-to-size sample design. Every establishment listed as belonging to the universe of entities producing in the to-be-sampled NAICS industry, regardless of size, has a chance of being selected. The chance of any single establishment's being chosen for participation in the survey is commensurate with its importance to the industry as a whole. Comprehensive coverage is necessary to insure that the price data collected is a representative sample of the universe of pricing activity within an industry. It is the PPI's opinion that the burden imposed on business establishments is very near the practical minimum consistent with production of a statistically meaningful index. (Survey References: Hill, Kimberley Dailey. (1987) "Survey Design in the Producer Price Index," ASA Proceedings of the Section on Survey Research Methods. Kulpinski, Stanley; Cohen Stuart J.; Perez-Lopez Kathleen, "Survey methods and theory of the Producer Price Index revision," ASA Proceedings of the Section on Survey Research Methods.) Additional documents are listed in Section C. PPI Methodology References.

    The steps involved in probability-proportional-to-size sampling include: constructing a frame (a list of businesses from which a sample is to be selected), identifying any specific variables that represent unique price-forming groups (explicit stratification), calculating the number of sample units and price quotations required within each unique group, sorting each group by a measure of size (usually employment), and using a calculated sample interval to select a representative subset of entities from the list. Probability-proportional-to-size sampling, in addition to improving efficiency and reducing bias, provides the capability to calculate statistical estimates of reliability, precision, and error.

    The number of establishments and price quotations selected for tracking varies, depending on the degree of homogeneity within the sampled industry. The sample must be large enough to represent the full range of producers and products. Since participation in the survey is voluntary, not every entity selected for inclusion cooperates.

Furthermore, sample frames typically contain a certain degree of error. Frame error includes entities defined as out of business and those incorrectly classified. Anticipated respondent attrition over the life of the sample also influences sample allocation.

Once a respondent has been approached by a BLS data collector and agrees to cooperate, initiation into the PPI survey requires, on average, 2 hours of respondent time. The amount of time varies, however, depending on the manner in which company records are kept, the number of price quotations requested, and the nature of the industry and products being sampled. The first step in initiating an establishment into the PPI involves verification of address and employment information. The next step involves identifying product lines produced or service lines provided, along with revenue data for each activity. The third step is item selection, which BLS refers to as disaggregation.

For each line of activity, respondents identify unique price-determining characteristics that come into play, along with the revenue that each line generates. A random number table is used to choose the unique transactions that will be tracked by PPI. This process is repeated for increasingly detailed categories until completely unique transaction types are identified. Disaggregation identifies unique price-determining variables, both product and transaction specific, and assigns a weighted importance to each. Identifying unique activities and their importance relative to the respondent's full revenue-generating activity allows the PPI to efficiently sample a representative subset of transactions, and permits efficient recordation of these classification parameters for future tracking. The BLS National Office provides forms to data collectors to assist in the process of assigning probabilities, selecting transactions, and documenting sampled transactions. (See attachments: forms BLS-1810A, BLS-18A1, BLS-1810-B, BLS-1810C, BLS-1810-C1, and BLS-1810E.)

Effective with the release of data for January 2004, the PPI converted its sampling, data collection, and industry-based publication structures to the NAICS. Through December 2003 PPI's industry-based procedures were linked to the SIC organizational system.

During monthly repricing, the main communication tool between the PPI and respondents is price-collection form BLS 473P. (See attachment.) The Program currently sends out approximately 105,000 forms per month to roughly 26,250 responding establishments. One form exists for each price quotation that is being monitored. This document contains the specific information required by the PPI to track changes in net transaction prices for predetermined outputs. Survey forms are designed to take industry-specific factors into account, allowing adaptation to individual company accounting and data structures. The program continues to evaluate the form in hope of streamlining and simplifying layout and content.

When price-quotation questionnaires are returned, they are entered into a database using an optical scanner. Respondents often submit forms that include changes to product descriptors, transaction descriptors, or net transaction prices. These changes may require a telephone call from a PPI industry analyst for clarification and verification.

Detailed-level price indexes are constructed by combining price quotations from respondents that describe similar product or service categories. Aggregate indexes -- whether they are product line, industry, industry group, commodity group, or stage-of-processing -- are weighted averages of detailed-level price indexes.

The modified Laspeyres formula provided below approximates the actual computation procedure for Producer Price Index:

$$I_t = \left[\left\{ \sum_{j=1}^{J} Q_{a,j} P_{0,j} \left(P_{t,j} \Big/ P_{0,j} \right) \right\} \Big/ \left\{ \sum_{j=1}^{J} Q_{a,j} P_{0,j} \left(P_{t-1,j} \Big/ P_{0,j} \right) \right\} \right] \times I_{t-1} \quad \text{(B.2.1)}$$

where P_o is the price of a product in the comparison period, P_t is the current price, and Q_a represents the quantity shipped during the weight-base period and the summation in this formula is over the index j which represents distinct products or groups of products and J is the total number of products or groups within a larger category. In this form, an index is the weighted average of price ratios for each item (P_t / P_o) in a detailed cell. The expression ($Q_a P_o$) represents the weights in dollar value form. The P and Q elements (both of which originally related to period a, but are adjusted for price change to period o) are not derived separately. When specifications or samples change, the index must be computed by linking (multiplying) the relatives for the separate periods for which the data are precisely comparable.

Within each PPI detailed cell, individual price quotation reports from establishments are given different weights, in accordance with data on shipment values provided to BLS field representatives by respondents during initiation interviews, adjusted by BLS using probability selection techniques. (Reference: Sager, Scott D. "Effect of Weights on Producer Price Indexes," *Monthly Labor Review*, July 1996.)

If a price quotation report has not been received in a particular month, then the change for that price will in general be estimated by mean imputation of price relatives within cells, i.e. averaging the price changes for the other items within the same detailed cell (that is, for the same kind of products) for which price reports have been received.

Item data are used in index calculation, and each item has its associated weight which is referred to as its item weight. An item weight has several components, which are related to item sampling weight, establishment sampling weight, and revenue.

To specifically understand how item weights are derived, one must consider the ramifications of the two-stage sampling process by which the items are selected. First, items which are selected in the second-stage estimate the first-stage sample unit, (i.e. establishment). The selected items estimate price movement for the establishment, and the item sample weights at the second stage sum to the establishment's total collected shipments and receipts at the first stage. For this reason, the establishment's collected shipments and receipts are considered only in association with the item sampling weight, and not with the establishment sampling weight.

The **formula for an item weight** is:

$$\text{Item weight} = \left[\frac{1}{\text{Reporting unit's probability of selection}}\right] \times$$

$$\left[\frac{\text{Product category percent}}{\text{Number of items attempted in category}}\right] \times \left[\text{Multiple hit factor of item}\right] \times$$

$$\left[\text{Reporting unit's shipments and receipts}\right] \quad (B.2.2)$$

where "product category percent" equals the proportion of a product category; "number of items attempted in category" equals the number of items within the selected product category that the field representative was instructed to collect; "multiple hit factor of item" equals the number of times the random-sampling process led to selection of this specific item; and "reporting unit's shipments and receipts" equals the total dollar value of the reporting units shipments and receipts in period a in (B.2.1).

Note that the first term is the reporting unit's first-stage sample weight, and the product of the last three terms equals the second-stage sample weight for the item.

It is important to note that in index calculation, an item weight determines the importance of its associated sampled item only *within the same product cell.* Even though some items may have roughly equal item weights, their importance to an aggregate index can be quite different if the items are in different product cells. If one ignores the effect of any price changes, then the importance that two items in two different product cells have relative to one another and to the aggregate index is a function of the Census weights (industry value of shipments data taken from the Economic Census) for the two cells and the relative item weight of the items within their respective cells.

3. Four months after first publishing its set of monthly indexes, PPI recalculates and finalizes indexes, taking into account late reports and back-corrections received from respondents. At this four-month mark, approximately 70% to 75% of price-quotation questionnaires are returned. Additional details on response rates are provided below. In order to maintain and improve cooperation, the PPI maintains a procedure that includes contacting, by telephone, any selected respondents that have not returned forms for a specified period of time. Assistance is provided with regard to any aspects of the form that at first glance appear unclear or burdensome; a common reason for nonresponse.

Response Rates for the PPI

Response Rates at Initiation

The following Response rates for the Producer Price Index (PPI) are for the Data Collection phase and are computed based on response at sample initiation of eligible

units (establishments) classified in various NAICS groupings. The current rates presented below were calculated for the NAICS structure using data collected and transmitted up to Feb 5, 2008. The data used for each industry was for the most recent sample for that industry. For industries not sampled under NAICS, SIC response data was mapped to the NAICS structure. Collection is complete for 403 industries that were sampled under NAICS.

Each response rate is an unweighted rate which is calculated by dividing the number of productive units (establishments) collected in a given grouping of industries by the total number of eligible units and units with eligibility undetermined in the same grouping.

As of Feb 2008, there were 22,209 eligible units in 495 Mining and Manufacturing industries for which the PPI publishes indexes and 14,271 eligible units in 149 Services industries for a total of 36480 eligible units. The establishment response rates for the above categories are given below. (Slack, David and Hagemeier, Kirk (2007) Survey Response Measurement Team Quarterly Report.)

PPI Response Rates at Initiation of Sample Units

		Mining and Manufacturing	Services	All NAICS codes
Feb	2008	85.2 %	78.9%	82.7%
Aug	2007	85.1 %	78.6%	82.6%
May	2007	85.1 %	78.6%	82.6%
Feb	2007	84.9 %	78.5%	82.4%
Nov	2006	84.6 %	77.9%	82.1%

Rates for Services industries grouped by NAICS sector have been calculated for Feb 2008 and are shown below. The Utilities sector has been included in Services.

PPI Response Rates at Initiation of Sample Units in Services Industries

Feb 2008

Wholesale Trade
        80.1%
Retail Trade
        76.2%
Transportation and Warehousing   86.2%
Information
        83.5%
Finance and Insurance   74.4%
Real Estate and Rental and Leasing   77.5%
Professional, Scientific, and Technical Services   71.6%
Administrative and Support Services[1]   71.6%
Educational Services   93.0%
Arts, Entertainment, and Recreation   94.7%
Health Care and Social Assistance   80.7%
Accommodation and Food Services   91.4%
Other Services excluding Public Administration   82.5%
Utilities
                87.8%

Response Rates at Repricing
The unweighted repricing response rate is calculated as follows:

All items used in estimation + All items with reported data that were not used in estimation) / (All items ever in estimation – (Discontinued Items + Out of season items + Off cycle items without a good price)). The unweighted item level repricing response rates are:

Feb.   2007   74.48%
Mar.   2007   73.68%
Apr.   2007   67.99%
May   2007   73.15%
Jun.   2007   71.64%
Jul.   2007   60.22%
Aug.   2007   68.37%
Sept.   2007   67.47%

Note that the response rates presented above are all unweighted rates. Due to the complex multistage sampling design and the complex multistage calculation of price

---

[1] The full title for "Administration and Support Services" is "Administrative and Support and Waste Management and Remediation Services".

indices, appropriate computation of weighted response rates is a topic of ongoing research, and will be considered further in future work.

The PPI will conduct an analysis of the non-response bias in its published data beginning in FY09. The PPI plans to focus the study on the two primary stages of data collection (initiation of primary sample units and monthly repricing of reported data). Since re-initiation is not a feasible alternative, the initiation phase of the study will be conducted by making comparisons between respondents and non-respondents across subgroups which can be identified using sample frame variables. When available, external sources of data will also be used. The repricing phase of the study will model response rate to survey variables at the item level to determine if non-response bias exists. In those cases where the analysis indicates that the data are not missing at random, the PPI will compare estimates produced through, respectively (a.) current weighting and imputation procedures and (b.) alternative weighing and imputation procedures based on the abovementioned response rate models.

4. The PPI is not currently planning any procedural or methods tests requiring OMB approval.

5. Oversight of statistical methods in the PPI survey are maintained by the Bureau of Labor Statistics, Office of Prices and Living Conditions, Division of Price Statistical Methods, Robert M. Eddy, Supervisory Mathematical Statistician, (202) 691-6932.

C. PPI Methodology References

The methodology of the PPI has been documented in numerous papers and articles written since 1977 when the PPI underwent the most comprehensive redesign in its history. These papers cover a broad spectrum of topics ranging from price theory and program concepts to actual data collection methodology. A list of references includes:

Archibald, Robert B. "On the Theory of Industrial Price Measurement: Output Price Indexes," *Annals of Economic and Social Measurement,* Winter 1977.

Bureau of Labor Statistics, *BLS Handbook of Methods,* U.S. Department of Labor. Available at http://www.bls.gov/opub/hom/homtoc.htm Chapter 14

Buszuwski, J.A. and Scott, S. (1988), "On the Use of Intervention Analysis in Seasonal Adjustment," *Proceedings of the Business and Economics Section,* American Statistical Association.

Buszuwski, J.A., "Alternative ARIMA forecasting horizons when seasonally adjusting producer price index data with X-11 ARIMA in concurrent mode" ASA Proceedings of the Business and Economic Statistics Section.

Collia, Demetra. (1988) "Measuring sample variability in the producer price index," ASA Proceedings of the Section on Survey Research Methods.

Council on Wage and Price Stability, *The Wholesale Price Index,* June 1977.
Early, John F. "Improving the Measurement of Producer Price Change," *Monthly Labor Review,* April 1978.

Gerduk, Irwin. (1984) "Quality assurance elements in Producer Price Index data initiation," ASA Proceedings of the Section on Survey Research Methods pp 151-156.

Hellerstein, Judith. (1989) "The effects of sample size on variances of the Producer Price Index," ASA Proceedings of the Section on Survey Research Methods pp 170-175.

Hill, Kimberley Dailey. (1987) "Survey Design in the Producer Price Index," ASA Proceedings of the Section on Survey Research Methods pp 583-588.

Kulpinski, Stanley; Cohen Stuart J.; Perez-Lopez Kathleen, (1978) "Survey methods and theory of the Producer Price Index revision," ASA Proceedings of the Section on Survey Research Methods pp 517-521.

Popkin, Joel. "Integration of a System of Price and Quantity Statistics with Data on Related Variables," *Review of Income and Wealth,* March 1978 pp 25-39.

Sager, Scott D. "Effect of 1992 Weights on Producer Price Indexes," *Monthly Labor Review,* July 1996 pp 13-23.

Sinclair, James and Catron, Brian. "New Price Index for the Computer Industry," *Monthly Labor Review,* October 1990.

Slack, David and Hagemeier, Kirk (2007) Survey Response Measurement Team Quarterly Report.

Spease, Carol . (1989) "Comparison of variance estimators for Producer Price Index data," ASA Proceedings of the Section on Survey Research Methods pp 176-181.

Tibbetts, Thomas R. "An Industrial Price Measurement Structure: The Universe Matrix of Producers and Products," *1978 Proceedings of the Section on Survey Research Methods.* American Statistical Association, Washington, DC, 1979 pp 511-516.

U.S. Department of Labor, Bureau of Labor Statistics, *BLS Handbook of Methods*. 2003. Chapter 14. Available at http://www.bls.gov/opub/hom/homtoc.htm

U.S. Department of Labor, Bureau of Labor Statistics, *Escalation and Producer Price Indexes: A Guide for Contracting Parties,* Report 807. September 1991.

U.S. Department of Labor Bureau of Labor Statistics, *Supplement to Producer Price Indexes, Data for 1990*. August 1991.

The Bureau of Labor Statistics in the Monthly Labor Review has published additional articles on specific PPI topics. A list of those articles includes:

- Consumer gasoline prices: an empirical investigation (July 2003)
- An empirical analysis of price transmission by stage of processing (November 2002)
- Producer price highlights, 2001 (July 2002)
- Producer price highlights, 2000 (July 2001)
- Producer price highlights, 1999 (August 2000)
- Producer price highlights, 1998 (July 1999)
- Comparing PPI energy indexes to alternative data sources (December 1998)
- The 1996 grain price shock: how did it affect food inflation? (August 1998)
- Producer price highlights, 1997 (July 1998)
- Improving the PPI samples for prescription pharmaceuticals (October 1997)
- Producer price highlights, 1996 (July 1997)
- Producer price highlights, 1995 (July 1996)
- Effect of 1992 weights on Producer Price Indexes (July 1996)
- Hospital price inflation: what does the new PPI tell us? (July 1996)
- Producer price highlights, 1994 (July 1995)

- Pricing practices for tobacco products, 1980-94 (December 1994)
- Producer price rises slowed in improving economy in 1993 (May 1994)
- Effect of updated weights on Producer Price Indexes (March 1993)
- Recession and energy prices ease producer prices in 1991 (May 1992)
- New price index for the computer industry (October 1990)