# SURVEY DESIGN IN THE PRODUCER PRICE INDEX

Kimberley Dailey Hill, Bureau of Labor Statistics
600 E Street NW Room 5217, Washington, D.C. 20212

BACKGROUND INFORMATION:

First published in 1902 as the Wholesale Price Index (WPI) and now known as the Producer Price Index (PPI), this survey is the oldest continuous statistical series published by the Bureau of Labor Statistics (BLS). In 1978 the survey was redesigned with new theoretical underpinnings and a new methodology. A comprehensive revision replaced the commodity-based indexes of the WPI with the industry-based structure in which indexes are based on the Standard Industrial Classification (SIC) scheme as developed by the Office of Management and Budget (OMB). The entire U.S. economy is divided into ten sectors. Current Producer Price Indexes cover all of the Mining and Manufacturing (M & M) sectors with ongoing expansion into the Service-related sectors.

There are 493 SIC classifications (in the 1972 manual) included within the scope of the Mining and Manufacturing sectors of the economy. The BLS is currently publishing indexes for nearly all of these industries. Currently there are more than 75,000 price quotes in repricing, representing approximately 18,000 reporters. The BLS has begun sampling some of these industries for the second time in the process of replacing outdated samples. The first time that each industry was sampled under the revised methodology is referred to as Cycle I, with Cycle II denoting the second time that an SIC is sampled. Cycle I was basically completed before Cycle II began, but since the frequency with which an industry is resampled is dependent on many variables, at some point in time, one industry will have been sampled three times, while another industry will have been sampled only once.

OBJECTIVES:

The primary objective of the 1978 revision, to produce indexes for all industries included in the Mining and Manufacturing sectors, was accomplished in early 1985. The industry-based structure implemented under the revision conforms to other Government statistical series. Also producing an index based on an industry structure eliminated the multiple counting of price change associated with the aggregate commodity indexes. Multiple counting had resulted in aggregate rates of price change being distorted unless the prices of all of the items at each stage of processing were changing at the same rate.[1] The revision also represented a move away from the judgmentally selected samples of the WPI to the probability samples of the current program. Experience gained during Cycle I has been used to further improve the procedures for succeeding cycles.

DETAILS OF THE SAMPLE DESIGN:

1. THE FRAME

The ideal frame would include all products produced by establishments classified in industries within the Mining and Manufacturing sectors of the economy. Such a list is not available. However, what is available is the Unemployment Insurance (UI) file, a list of all establishments, with each establishment classified in one and only one industry based on the plurality of its revenue. By law, every employer in the U.S. is required to report the number of people employed and to purchase insurance which will cover the employer's unemployment benefit liability. As a result the UI file data are fairly complete. The continued existence of the UI file is also assured, thereby ensuring continued availability of a consistent frame for sampling. The UI file contains such information as the establishment's name, the SIC in which it is classified, the county and state in which it is located, and its number of employees. This file is explicitly stratified according to industry classification and thus provides individual industry frames which form the basis for the PPI frames. Additional data sources are used to discover possible company affiliations.

The primary sampling unit for the PPI is the Profit Maximizing Center (PMC) referring to the economic unit within which prices are set, records are kept, and profits are maximized. The PMC may be either a single establishment or a group or "cluster" of establishments (operating within the same industry). The ultimate sample unit is an item selected in the sampled PMCs. Prices for these items, reported monthly, form the basis for the PPI indexes.

In Cycle I, the UI file was subdivided among the M & M industries, forming 493 mutually exclusive and exhaustive sampling frames. Though the UI file was updated annually, the SIC classification of a PMC was allowed to change only if the PMC had not previously had a chance of selection. Operationally, every frame cannot be sampled at the same point in time and since the classification of some PMCs may have changed, the resulting frames did not adequately reflect the industry existing at the time of sample selection. This situation in Cycle I was due mainly to the difficulty of incorporating known changes to the UI file data. In Cycle II this problem has been solved through the use of a Universe Maintenance System (UMS). Industry sampling frames are now formed to obtain an up-to-date frame for each industry at the time it is sampled. Therefore, an establishment may be allowed a chance of selection in more than one frame if the establishment has been reclassified into a different industry, either due to a frame error or to a shift in the production activities of the establishment. If it is determined that the establishment is producing in an SIC other than the SIC for which it is currently in repricing, the establishment will be removed from repricing and given a chance of selection in its correct SIC.

The UMS provides a systematic method for updating and correcting frames on a continual basis, so that the frames used for sampling are as up-to-date and accurate as possible. The information used to update the database comes from the annual UI files, as well as information gathered during field initiation, and during the course of other PPI processes. The UMS is used to update individual records and record charac-

teristics. It is also used to delete records that are currently out of business or that are outside the scope of the program. Using the UMS, records may also be added to the base file representing establishments which began production after the UI file was compiled for that particular year. One major advantage of the UMS is that an establishment which is misclassified can be reclassified into its correct SIC when the information is first discovered. A misclassified establishment is an establishment which is incorrectly classified into an SIC other than the SIC in which it has the plurality of its revenue.

Prior to the sample being drawn the UMS industry sampling frame is refined to improve its accuracy. During this stage clusters are formed containing the establishments believed to belong to particular PMCs. The larger frame units are contacted by telephone in order to determine the record-keeping practices of the companies, thereby confirming cluster formation. Additions or births to the frame or possible misclassifications may also be identified at this time. The final sampling frame is thus a modified version of the UI file satisfying the statistical and economic concerns of the PPI.

## 2. SAMPLE SELECTION PROCESS: STAGE 1

The sampling scheme employed by the PPI is a two-stage systematic probability proportional to size strategy. Two stages are required because a frame of all of the products produced in the M & M industries is not available. As was previously stated, the first stage sample refers to the sample of PMCs classified in the industry. Within each sampled PMC, a sample is drawn from all of the items produced by the PMC, regardless of whether an item is a primary product of the SIC being sampled. This constitutes the second stage sample.

Section 1 above describes the construction of the frame of all the PMCs classified in the SIC being sampled. At this time the frame may be explicitly stratified in an attempt to achieve strata within which price movement and relevant economic factors tend towards homogeneity. One example of this type of stratification would be by geographical location. This stratification can contribute to the precision of the estimated index, as well as allow more accurate estimation of missing revenue values. When information is not available with which to classify the PMCs into the proper strata, post stratification may be used, based on information collected when the PMC is initiated. If homogeneous strata cannot be formed then the entire industry frame is viewed as a single explicit strata.

After the frame has been explicitly stratified, the units are ordered within each stratum. It is expected that the establishments which are close to each other in the ordered list will tend to have similar pricing trends. The ordering is based on either employment size or geographic location of the establishment depending on which variable is believed to have a greater effect on price movement. Systematic sampling within each stratum then results in a forced spread of sample units with respect to this ordering.

In the PPI, the ideal measure of size would be the PMC's revenue. The UI file does not contain this information, but does include the number of employees. This is used as a proxy in sampling since the number of employees tends to be correlated with the revenue of a PMC within a particular SIC. In the PPI, a PMC having employment larger than the sampling interval is selected as a certainty unit. The estimation implications of each of these cases is discussed in Section 5 below.

Early in Cycle I a change was made in the sample design in order to permit the computation of variances using balanced half-sample replication. Previously the sample was one systematic probability proportional to size sample within each explicit stratum. Currently within each explicit stratum, the frame is partitioned into segments of approximately the same size, called variance strata. The sample of PMCs is drawn by taking two independent systematic probability proportional to size half-samples within each variance stratum. The number of variance strata formed is a function of the number of replicates needed to calculate variances and the quantity of data available. A replicate is a subset of the total sample. Each replicate is formed by taking a specified combination of the half-samples from each of the variance strata. The remaining data are denoted as the complement of the replicate. In the PPI, each industry will have either 4, 8, 16, or 32 replicates and complements assigned to it. An index is calculated at every level of detail for each replicate and complement using exactly the same methodology as for the overall index. Imputations for missing data are computed independently within each replicate and complement. The variance of the index is computed using the following formula:

$$\hat{V}_I = \sum_{q=1}^{Q} \frac{(\hat{I}_{q,r} - \hat{I}_{q,c})^2}{4 \times Q}$$

where $\hat{V}_I$ = variance of industry I

$Q$ = number of replicates

$\hat{I}_{q,r}$ = index calculated using data from the $q^{th}$ replicate

$\hat{I}_{q,c}$ = index calculated using data from the $q^{th}$ replicate's complement.

## 3. SAMPLE SELECTION PROCESS: STAGE 2

The second stage sampling technique used by the PPI is also a systematic probability proportional to size scheme. This technique is referred to as disaggregation. Basically the procedure involves having the respondent classify all products produced by the sampled PMC into broad categories based on the PMC's records. Product categories are then selected (according to the assigned quote allocation) employing a probability proportional to size systematic sample. An attempt is made to use the actual percentage of the PMC's dollar value of revenue for each product category as the measure of size. If the actual percentages are not available then estimates are used. These estimates are based on respondent knowledge, ranking with

assigned probabilities, or equal probability (in order of preference). Within each selected category, the process is repeated using more detailed product descriptors until the process identifies a unique item. By definition of being unique, the item has no other price-determining characteristics which would call for further disaggregation.

As with the first stage sampling method, the second stage procedure was also altered by the decision to support variances. For PMCs that are chosen in both half-samples as well as for PMCs which are selected as certainty units, the number of quotes assigned to the particular schedule is divided between the two half-samples and disaggregation is performed independently in each.

Normally, between three and eight price quotes are allocated to each sampled PMC. The number is based on the number of product categories for which publication is desired, as well as on whether the PMC is in both half-samples or is a certainty unit. Other considerations can require adjustments to the allocations. For instance, substantial variability in type of production or price movement within the PMCs in the industry requires a larger quote allocation for each PMC. Additionally, special first and second stage strategies are used to reduce reporting burden for the largest companies.

## 4. ADDITIONAL CYCLE II REFINEMENTS:

A study has been made of the disaggregation performed during Cycle I. An attempt is now being made to define on a PMC basis when sufficient disaggregation has been performed to ensure the proper spread of quotes and weight across the industry. The aforementioned study indicates that this definition will vary by industry, due to the differing economic factors within each industry. At present, the information currently requested during the initiation of the PMC is being revised to support further work in this area.

In an attempt to tailor the sample design to the specific situation, two industries, pharmaceuticals and paper mills, were collected using an altered second stage sampling technique. It was clear that the sample allocation was too small to support the publication goals using standard second stage sampling techniques. During industry analysis, these industries were discovered to be very diverse, having many detailed product categories. The larger PMCs, mirroring the industry, produce a wide range of products, while the smaller PMCs have production in only a few categories. It was agreed that a quote allocation increase for the larger PMCs would be more beneficial in solving the problem of industry coverage than increasing the number of sample units. In the normal second stage pro-cess, certainties are not removed and multiple hits are allowed. Since the objective of this particular sample design was to obtain more complete coverage of detailed product categories than would normally occur, an alternative disaggregation scheme was developed. The procedure parallels the current first stage process in that certainties are removed, thereby reducing the number of multiple hits. Instead of allowing the company to define the disaggregation path, the primary disaggregation path was fitted to the detailed product categories of the particular industries. Using this procedure, the necessary spread of quotes across all product categories was achieved. This was an extension of procedures used on a very limited basis in Cycle I.

At the close of Cycle I, efforts were directed towards devising a timetable for resampling the M & M industries. Not every industry needs to be resampled at the same rate. A procedure has been developed that facilitates the decision about when to resample by standardizing the method for making this decision. The procedure takes into consideration such criteria as: technological change since the last time sampled, number of products for which substitutions have been made in the current sample, overall change in the number and composition of establishments in the base file, the response rate in the industry, publishability of the cells within the industry, quality of the current sample and any changes to the SIC manual which affect this industry.

## 5. ESTIMATION METHODOLOGY:

The PPI is a modified Laspeyres index. A Laspeyres index is a weighted average index comparing the prices of a group of items in any time period to the prices of those same items in the base period. In the economic concept underlying the PPI, certain simplifying assumptions can be made. First, it can be assumed that a PMC will work towards maximizing its profits in a perfectly competitive market. The types and quantities of inputs, and technology, are assumed to remain constant from the base period forward. Lastly, it can be assumed that the PMC's output mix remains fixed to that of the base period. Under these assumptions, in order to increase its profit, a PMC can only change its prices, and the PPI measures change in revenue from a base period due to changes in prices. In theory this means that only pure price change, not change due to quality differences, should be reflected in the indexes. If the assumption of fixed output ratios is relaxed, the Laspeyres index becomes a lower bound of the true economic output price index.

The general formula for a Laspeyres price index is of the form:

$$I^{t,b} = \frac{\sum_i \sum_j P_{ij}^{\,t} Q_{ij}^{\,b}}{\sum_i \sum_j P_{ij}^{\,b} Q_{ij}^{\,b}} \times 100 \qquad [1]$$

where: $I^{t,b}$ = Index of change from base period $b$ to time $t$

$P_{ij}^{\,t}$ = price of item $j$ in PMC $i$ at time $t$

$Q_{ij}^{\,b}$ = base period quantity of item $j$ in PMC $i$

$P_{ij}^{\,b}$ = base period price of item $j$ in PMC $i$

Multiplying by 100 allows the reference period index to equal 100. This formula assumes that every item in a given industry will be priced. Time and budget constraints make this impossible, therefore a sample of items is selected.

Development of the Estimator, $\hat{V}^t$:

Since it is known that ratio estimators are not unbiased, the best that can be shown is the unbiasedness of the numerator and denominator. The numerator in formula 1 can be expressed as follows.

$$V^t = \sum_{i=1}^{N} \sum_{j=1}^{M_i} V_{ij}^b \left(P_{ij}^t / P_{ij}^b\right) \qquad [2]$$

where $V^t$ is the total revenue at time $t$

$V_{ij}^b$ is the base period revenue for item $j$ in PMC $i$ $(=P_{ij}^b Q_{ij}^b)$

$N$ = number of PMCs

$M_i$ = number of products produced by PMC $i$

Now consider an estimator for $V^t$:

$$\hat{V}^t = \sum_{i=1}^{N} \frac{\gamma_i}{Pr(i \varepsilon S)} V_{i\cdot}^b \sum_{j=1}^{M_i} \frac{1}{m_i} \gamma_{ij} \frac{P_{ij}^t}{P_{ij}^b} \qquad [3]$$

where $S$ denotes the first stage sample

$\gamma_i$ is $\begin{cases} 1, & \text{if PMC } i \in \text{Sample } S \\ 0, & \text{otherwise} \end{cases}$

$\gamma_{ij}$ is the number of times item $j$ is selected in PMC $i$

$V_{i\cdot}^b$ = base period revenue for PMC $i$

$m_i$ = number of products selected in $i$

In the PPI, several factors related to the index estimator should be noted:

a. When selecting the first stage sample, employment is used as a proxy for revenue since the two tend to be correlated and the employment figure is available.

b. During disaggregation the value assigned to each product category may be the actual revenue value or could possibly be an estimate based on respondent knowledge.

Development of the probabilities used in the estimator, $\hat{V}^t$: As described in section 2 above, the first stage probabilities are based on the PMC's employment relative to the rest of the industry such that:

$$Pr(i \varepsilon S) = \begin{cases} 1, & \text{if PMC } i \text{ is a certainty selection} \\ 1 - \left(1 - \left(\frac{n_h e_i}{\sum e_i}\right)\right)^2, & \text{otherwise} \end{cases}$$

where $n_h$ is the number of non-certainty selections in each half-sample

$e_i$ is the UI sample employment for the PMC $i$

$\sum e_i$ is summed over all non-certainty PMCs.

The second stage of sampling consists of multiple steps of disaggregation, in which continually more detailed product categories are selected from the broader categories containing them. Assume there are $k$ steps of disaggregation performed. At each step, $p$, the probability of selecting a product category is $\frac{V_p^b}{V_{p-1}^b}$ for $p=1,\ldots,k$ where $V_0^b = V_{i\cdot}^b$ (the PMC revenue value) and $V_k^b = V_{ij}^b$ (the item revenue value).

Therefore, $E(\gamma_{ij}) = m_i \prod_{p=1}^{k} \frac{V_p^b}{V_{p-1}^b} = m_i \frac{V_{ij}^b}{V_{i\cdot}^b}$

where $m_i$ is the number of items selected in PMC $i$

Now substituting into formula 3 yields:

$$\hat{V}^t = \sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^b \sum_{j=1}^{M_i} \frac{\gamma_{ij}}{m_i} \frac{P_{ij}^t}{P_{ij}^b} \qquad [4]$$

where $a_i = \begin{cases} 1, & \text{if PMC } i \text{ is a certainty unit} \\ \dfrac{1}{1-\left(1-\left(\frac{n_h e_i}{\sum e_i}\right)\right)^2}, & \text{otherwise} \end{cases}$

In the second stage sample, items may be selected more than once within a PMC due to the selection procedure used in the two half-samples and the relative importance of the item in the PMC. This, in combination with the possibility of certainty selected items, imposes another restriction on the above estimator. The use of $V_{i\cdot}^*$, the revenue value for the year preceding initiation, instead of the base period revenue represents a departure from the theoretical Laspeyres concept. These values are benchmarked to Census values for the base year at the detailed product cell level. Thus, the actual estimator used for the PPI is:

$$\hat{V}^t = \sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{M_i} \gamma_{ij} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} \qquad [5]$$

where $V_{i\cdot}^*$ is the revenue for PMC $i$ at the time of initiation

$V_{ij}^*$ is the revenue attributed to item $j$ in PMC $i$

$$r_{ij} = \begin{cases} \frac{1}{m_i}, & \text{if there are no certainty selected items in PMC } i \\ \frac{V_{ij}^*}{V_{i\cdot}^*}, & \text{for each certainty selected item} \\ \dfrac{1 - \sum_{j=1}^{n_c} \frac{V_{ij}^*}{V_{i\cdot}^*}}{m_i - n_c}, & \text{for each non-certainty selection in a PMC containing certainties. The number of certainties is denoted } n_c. \end{cases}$$

Note that $\sum_{j=1}^{m_i} \gamma_{ij} r_{ij} = 1$ for each PMC $i$.

Taking the expected value of $\hat{V}^t$ as given in formula 5, yields the following:

$$E(\hat{V}^t) = E\left(\sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{M_i} \gamma_{ij} r_{ij} \frac{P_{ij}^t}{P_{ij}^b}\right)$$

Consider the formula for conditional expectations:

$$E(\hat{V}^t) = E(E(\hat{V}^t | \gamma_i \; i=1,\ldots,N))$$

Now consider

$$E(\hat{V}^t | \gamma_i \; i=1,\ldots,N) = [E(\sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{M_i} \gamma_{ij} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} | \gamma_i \; i=1,\ldots,N)]$$

$$= [\sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{M_i} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} E(\gamma_{ij} | \gamma_i \; i=1,\ldots,N)]$$

Without loss of generality, assume the first $n_c$ products within each PMC are selected with certainty.

$E(\hat{V}^t | \gamma_i \; i=1,\ldots,N)$ can then be expressed as:

$$[\sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{n_c} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} E(\gamma_{ij} | \gamma_i \; i=1,\ldots,N)$$

$$+ \sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=n_c+1}^{M_i} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} E(\gamma_{ij} | \gamma_i i=1,\ldots,N)]$$

For the first $n_c$ products selected with certainty

$$E(\gamma_{ij} | \gamma_i \; i=1,\ldots,N) = 1 \text{ and } r_{ij} = \frac{V_{ij}^*}{V_{i\cdot}^*}$$

For the remaining products

$$E(\gamma_{ij} | \gamma_i i=1,\ldots,N) = \frac{(m_i - n_c) V_{ij}^*}{V_{i\cdot}^*} \text{ and } r_{ij} = \frac{1}{m_i - n_c}$$

Substituting into the original equations yields:

$$E(\hat{V}^t | \gamma_i \; i=1,\ldots,N) = \sum_{i=1}^{N} \gamma_i a_i \sum_{j=1}^{n_c} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}$$

$$+ \sum_{i=1}^{N} \gamma_i a_i \sum_{j=n_c+1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}$$

$$E(\hat{V}^t | \gamma_i \; i=1,\ldots,N) = \sum_{i=1}^{N} \gamma_i a_i \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}$$

Taking the expectation over all PMC's results in:

$$E(E(\hat{V}^t | \gamma_i i=1,\ldots,N)) = E(\sum_{i=1}^{N} \gamma_i a_i \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b})$$

$$= \sum_{i=1}^{N} a_i \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b} E(\gamma_i)$$

Without loss of generality, assume the first k PMCs are selected with certainty. The equation can then be written as:

$$E(E(\hat{V}^t | \gamma_i i=1,\ldots,N)) = \sum_{i=1}^{k} a_i \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b} E(\gamma_i)$$

$$+ \sum_{i=k+1}^{N} a_i \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b} E(\gamma_i)$$

For the k PMCs selected with certainty, $E(\gamma_i)=1$ and $a_i = 1$. For the remaining PMCs:

$$E(\gamma_i) = 1-(1- \frac{n_h e_i}{\sum e_i})^2 \text{ and } a_i = \frac{1}{1-(1- \frac{n_h e_i}{\sum e_i})^2}$$

where $n_h$ denotes the number of probability units selected and the summation of the $e_i$ is over those units not selected with certainty.

Substituting into the equation yields:

$$E(E(\hat{V}^t | \gamma_i \; i=1,\ldots,N)) = [\sum_{i=1}^{k} \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}$$

$$+ \sum_{i=k+1}^{N} \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}] = \sum_{i=1}^{N} \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}$$

Therefore $E(\hat{V}^t) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} V_{ij}^* \frac{P_{ij}^t}{P_{ij}^b}$

This completes the proof that $\hat{V}^t$ is an unbiased estimator of the numerator in formula 1. The estimator for the denominator ($V^b$) differs from that for the numerator only in that it lacks the (statistical) constant $P_{ij}^t/P_{ij}^b$. Therefore the preceding proof also demonstrates the unbiasedness of the demoninator,

$$\hat{V}^b = \sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{M_i} \gamma_{ij} r_{ij}$$

In the PPI, an estimate of price change must be produced for each cell in the aggregation structure. In order to meet this requirement a Laspeyres index must be generated at this level of detail. The formula used to generate an estimate of the numerator is:

$$\hat{V}_c^t = \sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^* \sum_{j=1}^{M_i} \gamma_{ij} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} \alpha_c$$

where $\hat{V}_c^t$ = total revenue for cell c at time t

$\alpha_c$ 1 if item j is in cell c

0 otherwise

Since $\alpha_c$ is not a function of the selected sample but simply restricts the set of items, a proof exactly like the one above yields:

$$E(\hat{V}_c^t) = V_c^t$$

Thus, the index estimator used by the PPI is:

$$\hat{I}_c^t = \frac{\hat{V}_c^t}{\hat{V}_c^b}$$

$$= \frac{\sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^{*} \sum_{j=1}^{M_i} \gamma_{ij} r_{ij} \frac{P_{ij}^t}{P_{ij}^b} \alpha_c}{\sum_{i=1}^{N} \gamma_i a_i V_{i\cdot}^{*} \sum_{j=1}^{M_i} \gamma_{ij} r_{ij} \alpha_c} \times 100 \qquad [6]$$

Weighted aggregation is the method used to combine the various detailed product cells in an industry into more general, or higher level, product categories. The manner in which a particular industry is aggregated is determined by the publication structure of the SIC. Using weights derived from Census data, these lower level cell indexes are combined to produce an index for the higher level product categories. This aggregation process is continued until an index is calculated for the industry. The PPI has a four month correction policy allowing prices to be revised up to four months after the index was calculated. To accommodate this correction policy, the cell indexes for the previous four months are recalculated each month. The cell index for the current month is calculated by chaining these recalculated cell indexes together. Similarly, due to the changes in the weights caused by the addition or deletion of cells, the aggregate index is calculated in the same manner.

One problem encountered when collecting data, is either a partial or total refusal on the part of the respondent to provide data. The PPI program uses Census data for revenue values at the lowest level cells, thereby benchmarking the data. In this manner, each cell carries its proper weight and any refusal weight will be moved by respondents for its cell. Since price movement is assumed to be most similar within a lowest level cell, it is more reasonable to reweight for any nonresponse within the lowest level cell than to increase the weight across the entire industry.

Similar to the problem of nonresponse, and handled similarly, is that of estimating missing prices. In this case, the respondent fails to report the price of a selected item for one or more months. Since index calculation in the PPI is based on monthly price change, a consistent method for estimating any missing prices is necessary. The cell relative method is generally used in PPI calculations to estimate prices for which data were not received. The missing price is estimated by multiplying the price of the item in the previous month by the cell's average price change in the current month. In this way, the missing price is estimated based on the prices of other items contained in the same cell.

If an item is not being repriced due to the fact that the item is no longer being produced, then a substitution is made. The item chosen as a replacement should be similar in function to the discontinued item. The item may be deleted from the sample if an appropriate substitute cannot be found.

When the respondent fails to provide revenue data, this figure must also be estimated. This dollar value is estimated based on a calculated average dollar value per current employee for the reporters in the same stratum as the PMC with the missing revenue value.

Stage of Processing (SOP) indexes are also calculated using the data collected to produce the industry indexes described above. These basically divide the industries of the PPI into four categories; crude, primary, semi-finished, and finished products. These indexes show how price change moves through the different stages of the economy. The main objective is to assign the industries to a stage in such a manner as to maximize revenues flowing from one stage of processing to the next stage forward. Further, SOP indexes attempt to minimize the flow of products that skip at least one stage, say from crude to finished products.

6. FUTURE PLANS:

Prior to 1978, the major PPI indexes were organized by commodity groups. Commodity indexes are still being published by regrouping data collected for the PPI industry indexes. The index for a commodity is based on price data collected only in the industry for which the commodity is primary production. In the future, the BLS plans to publish wherever-made indexes. These indexes will be similar to their predecessors except that they will be structured using all price data for a given commodity, regardless of the industry in which the commodity was produced.

Work is currently being done to assess the affects of the 1987 SIC Manual Revision on the PPI. This revision represents the first substantial changes to the manual since 1972. Industry definitional changes have been made to approximately one-third of the 493 industries included within the M & M sectors. Unfortunately even the most simple change may have serious implications for the systems used to store and process the data as well as provide for the continuity of the indexes. The primary non-operational problem caused by the SIC revision is the desired, but unobtainable ability to historically recalculate indexes using the new industry definitions. The impact of this revision on the Cycle II industries currently in publication is being studied to determine the most efficient plan for handling these changes.

Service industries account for a very large percentage of the revenues in the U.S. economy. The BLS is currently publishing a few service industries, at varying levels of detail, but basically the services initiative is in its early stages. Examples of service industries in publication include scrap and waste materials, railroad transportation, telephone communications and oil pipelines. Given current budget constraints, three to six service industries are expected to be sampled in the coming year.

Other aspects of the PPI are being reviewed. Plans are being made to conduct a user's survey to determine possible alterations to current data and any additional data requests. The PPI is continuously being revised in an attempt to accurately measure price change in the ever-changing primary markets of the U.S. economy. It is hoped that the indexes will be improved by reflecting previous cycle experience and user need.

Footnotes

[1] Gousen, Sarah and Monk, Kathy, PRODUCER PRICE MEASUREMENT - CONCEPTS AND METHODS p. 15-18