SURVEYING INTERNATIONAL PRICES

Marvin Kasper and Richard J. Pratt, Bureau of Labor Statistics

Introduction

The International Price Program produces export and import price indexes to measure price change. Export or import indexes, as all Bureau of Labor Statistics (BLS) price indexes, are Laspeyres indexes which measure price change over time of fixed quantities and sets of items. To support these indexes, a survey had to be designed to provide a sample of exporters and importers as well as specific items in each which could be repeatedly priced over time. Indexes are published for classes of items; therefore, the survey design had to support each of these commodity class indexes.

Commodities imported or exported by an establishment may change as international trade conditions change. For example, a change in the exchange rate may cause an establishment to change the products in which it trades or perhaps shift from imports to exports. This is particularly true of importers who are buyers and can more readily change the commodities in which they trade; therefore, an import price index is vulnerable to changes in international trade conditions. To produce price indexes the survey design had to address this lack of periodicity of items in establishments.

The survey design also had to be responsive to cost constraints. These constraints imposed limits on the number of distinct establishments that could be selected for the sample. In addition, the number of items priced in each establishment had to be controlled to limit respondent burden. The survey design attempts to address all of the above constraints while maintaining a uniform work flow compatible with available field resources.

Requirements

To illustrate how program requirements are implemented in the survey design a brief description of the export and import product classification systems will be given. Exports are classified in two seven digit systems, the Schedule B and the new 1978 Schedule E of the U.S. Department of Commerce. Correspondingly imports are also classified using two seven digit systems, the Tariff Schedule of the United States Annotated (TSUSA) and the Schedule A. The Schedule E for exports and Schedule A for imports are nested systems based on the Standard International Trade Classification (SITC), that is, a one digit product category is subdivided by adding a second digit which in turn is subdivided by adding a third digit and so on through the seven digits. The seven digit categories of the Schedule E are further subdivided using the seven digit categories of the Schedule B. The seven digit categories of the Schedule A are subdivided using seven digit categories of the TSUSA. When an exporter files a Shippers Export Declaration, he classifies and enters the items exported using the Schedule B classification. Correspondingly importers classify commodities on import documents using the seven digit TSUSA classification system.

To meet program objectives, the Bureau is in the process of expanding product areas for which indexes are published. The expansion is done in phases to maintain a uniform collection effort with limited resources. When expansion occurs, the Bureau specifies a section or set of subclasses for which indexes are to be published. The publishable classes can be at any level from the seven digit level to an overall one digit level index. Those classes containing no publishable subclasses are defined to be publishability strata. The items in a publishable class not contained in any publishable subclass are formed into strata called the residual strata of the given class. The publishability and residual strata partition the commodities into disjoint sets. Consider the following example:

> Publish indexes for categories 6, 62, 6291015 and 65.
>
> Publishability strata 6291015 and 65 (contain no publishable subclasses).
>
> Residual strata 62R (those commodities in 62 not in 6291015), 6R (those commodities in 6 not in 62 or 65).
>
> 6291015, 65, 62R, and 6R are disjoint and exhaust 6.

The publishability and residual strata are the strata used to construct the survey design.

The classification system is also used to define and construct entry level items (ELIs). These are commodity classes with which a field representative enters an establishment and then proceeds to obtain a specific item in this class which is priced over time. An "entry level item" contains homogeneous items with respect to price change. For exports, ELIs contain one or more seven digit B numbers within a stratum. For imports, the seven digit A number is the ELI.

Sampling Frame

To construct a sampling frame, a reference period, generally one year in length, is specified. All Shippers Export Declarations or Import Documents filed during the reference period which contain items in the required commodity classes are used as the data base for the sampling frame. Because of differences in export and import records, different techniques are used to refine the sampling frame.

Export Frame Preparation

When a product is exported, the exporter or his agent files a Shippers Export Declaration. Each entry line on the declaration corresponds to a seven digit Schedule B category. The exporter classifies the product and enters its Schedule B number and dollar value. The Schedule B has many

similar and complex categories, therefore, products could be misclassified. Because no tariffs are involved, Customs does little verification of the information reported on the declarations. An additional problem is caused by changes introduced into the Schedule B. Exporters not aware of changes may incorrectly classify products.

The Shippers Export Declaration is assigned a unique ten digit number designating the block, file, and month. The company name and address are given on the declaration. Since there are no export duties, a unique identification number has not been given to each exporter. The Census Bureau creates a computer file of all the information on the declaration excluding the name and address of the exporter. Each record on the file corresponds to a line on the Shippers Export Declaration. Since the name and address are not on the computer file, the paper file where the declarations are stored must be searched using the block, file, and month designator to obtain this information. In addition, because there is no unique exporter code, a matching process using man machine interface has to be used. These two problems limit the number of documents that can be processed; therefore, a double sampling technique is used with the initial sample of export declarations selected by Census.

The initial sample is selected by Census from the computer file with probability proportionate to size (PPS) using the dollar value of each line on the Shippers Export Declaration as the measure of size. The file is sorted by strata and within each stratum by entry level item. A total of 12,000 lines corresponding to at most 12,000 declarations are allocated. This number was chosen to meet program needs or to comply with resource restrictions. The allocation of lines to each stratum is done in two stages. An initial allotment is assigned to meet publishability requirements and the remainder are allocated to each stratum proportionate to the total dollar value of all the lines in the stratum. Units with measure of size greater than the sampling interval are designated as certainty units. A PPS selection is made from the remaining units. After this initial selection of lines all other lines in the required categories having the same block, file, and month as at least one of the selected lines are added to the sample, that is, all lines in the required categories on a declaration which had a line selected are in the sample. The resulting sample consists of approximately 12,000 export declarations containing 25,000-30,000 lines.

Using the block, file, and month (document identifier) the paper file of Shippers Export Declarations is then searched for the name and address which are added to the file of sample units. Approximately 93% of the sample declarations can be matched, that is, for 7% of the selected units a name and address cannot be obtained. Units may not be matchable because the document cannot be located or the name and address are not on the document or are illegible. The sample file with the added name, address, and sampling weight are then forwarded to the Bureau of Labor Statistics in machine readable form. The sample file is carefully refined using both the name and the address to obtain a unique exporter code and this code is added to each record.

Import Frame Preparation

While the basic construction of the import price index is similar to that for exports, there are differences in the universe which result in a different sampling approach. The most significant difference in the import universe is the availability of an identification number on the documents (and also on a computer file) which is unique to an importer. All importers are required by law to be registered with the Customs Bureau. If the ultimate consignee (the owner of the goods) is importing the merchandise himself, then he will be registered with Customs. If the company is not registered with Customs and is using a broker as his importing agent then the agent is required to file his own identification number along with registration material used in obtaining an identification number for the ultimate consignee. However, there are certain exceptions to this practice which result in a file called unmatched consignees. This unmatched universe has accounted for between 5% and 15% of the total dollar value in our previous samples. Two ways of handling this problem are: 1) a manual search similar to the one done by the Census Bureau in exports or 2) contacting the agent and trying to locate the ultimate consignee through him.

The file obtained from the U.S. Customs Bureau is a universe of all import documents during the reference period in the sections requested. The sections completed at this point have averaged approximately 1.5 million documents filed. The information retained from each document includes:

1) the consignee number
2) the agent number
3) the name and address of the consignee
4) the name and address of the agent
5) the dollar value of the transaction
6) the TSUSA under which it is filed
7) the month in which the import occurred.

An A-number and a stratum code are then added using a concordance file.

A verification procedure by Customs has already taken place in order to assure that proper payment of duties for each transaction has been made [1], [2]. The BLS checks the data by aggregating the dollar values up to the TSUSA and A-number levels and comparing these reference period totals to those published by the Census Bureau.

Some of the larger companies will have more than one establishment identifier. Since approximately 60,000 ultimate consignees are registered with Customs, the only matching procedure which is feasible is a computer check based on exact name. If a company is in the file

with more than one name (abbreviation, division, misspelling) then each entry is treated as a separate importer until after a sample of companies is drawn. The remainder of the discussion on the Importer Frame Preparation will concern the matched universe after combinations have been made.

The import market is a purchasers market; therefore, there is a significant amount of non-repriceability among the products. This can be seen in Table I where most of the trade (62.6% and 58.3%) is by companies trading only once in the 12 month reference period in the TSUSA. This is contrasted to the amount of imports transacted by companies trading every month of the reference period in a TSUSA (2.6% and 1.8%). The situation may be more severe since Table I shows company-TSUSA's and not company-products. This shifting of product lines in an establishment resulting in nonrepriceability is the program's most serious problem and because of this a tremendous amount of effort has gone into analyzing the data and trying to develop a workable solution.

TABLE I

Profile of Company–
TSUSA Distribution by Months of Active Trade

No. Months Traded in Reference Period	Company - TSUSA's 2nd Importer Sector No.	%	3rd Importer Sector No.	%
1	92,310	62.6	72,879	58.3
2	18,872	12.8	18,897	15.1
3	9,406	6.4	9,720	7.8
4	5,873	4.0	6,005	4.8
5	4,222	2.9	3,989	3.2
6	3,463	2.3	3,061	2.5
7	2,431	1.7	2,181	1.7
8	1,984	1.3	1,859	1.5
9	1,762	1.2	1,528	1.2
10	1,552	1.0	1,339	1.1
11	1,795	1.2	1,300	1.0
12	3,822	2.6	2,305	1.8
	147,492	100	125,063	100

Owners are identified as either consistent importers or inconsistent importers. In addition, the TSUSA's, A-number's, and strata within companies are classified into groups according to their periodicity. This is done as follows. A score is assigned to a company-TSUSA based on the product of 1) the total number of months in the reference period in which the owner imported in a given TSUSA and 2) the total number of quarters in the reference period in which the owner was active in that TSUSA. The score is then converted into a consistency rank for that owner-TSUSA. Once a consistency rank has been assigned to all owner-TSUSA's, consistency ranks can be assigned at the owner-Schedule A-number, owner-ELI, owner-stratum, and owner levels by using the maximum consistency rank over all ranks within the appropriate sub-level. Thus the rank for an owner-stratum would be the maximum rank of all owner-ELI's within that owner-stratum.

The consistency ranks will be used in both obtaining a sample of importers (first stage) and subsampling items within an importer (second stage). They will be used at the owner level in the first stage to stratify the owners into a consistent group and an inconsistent group. Consistency ranks at the owner-ELI level will be used in such a way as to give those ELI's with higher consistency ranks a greater chance of selection in the second stage selection. This is to increase the chance of choosing repriceable items. The program objective is to produce an index representing all imports. While practical considerations restrict us to frequently traded or repriceable items, all items are given a chance of selection and weighted accordingly.

The reasons for concentrating so heavily on consistency can be seen in Table II and Table III. In Table II it is shown that while 74% of all company-ELI's were inconsistent (class 1 or 2) these accounted for no more than 8.5% of the total dollar value. This is contrasted to classes 4 and 5 which accounted for at least 86.5% of the dollar value while making up no more than 20% of the total number of company-ELI's. Table III exhibits how the response rates in the survey were distributed by consistency class. As can be seen, the higher the consistency rank the better the cooperation rate. Also there is a 50% reduction in the out of scope rates when one goes from class 1 to class 5. There is no pattern concerning the refusal rate. There is a better return from the more consistent ELI's in terms of numbers, and also a heavy concentration of the dollar value covered by the Index.

TABLE II

Distribution of Imports
by Frequency of Trade in Establishments

2nd Importer Sector

	(infrequent)				(frequent)
Consistency Class	1	2	3	4	5
Percentage of total number of company-ELI's	61	13	6	11	9
Percentage of total dollar value	5	2.5	2.5	10	80

3rd Importer Sector

	(infrequent)				(frequent)
Consistency Class	1	2	3	4	5
Percentage of total number of company-ELI's	59	15	8	11	7
Percentage of total dollar value	5.5	3	5	14	72.5

Survey Design

The objective of the International Price Program is to produce a Laspeyres Index of import and export prices.

TABLE III

Entry Level Item Response Rates
by Frequency of Import

2nd Importer Sector

Consistency Rank	% Cooperation	% Refusal	% Out of Scope
1	7	12	81
2	15	15	70
3	20	17	63
4	22	19	59
5	37	23	40

3rd Importer Sector

Consistency Rank	% Cooperation	% Refusal	% Out of Scope
1	10	11	79
2	18	11	71
3	23	9	68
4	35	13	52
5	45	19	36

This essentially means estimating the updated dollar value of a fixed set of imports or exports. The design had to maximize the number of prices obtained from each company while maintaining control on the number of distinct companies in the sample. In addition the survey had to support indexes for each publishability class. To accomplish these goals, a two stage design which is based on the technique first presented by Lahiri [3] was implemented. The result is an estimator which is the sum of the dollar value in each stratum updated by the unweighted average of the price relatives (See Estimation) in the stratum. In addition, the expected number of sample units in each stratum is the same. Practical considerations have resulted in certain modifications which will be noted.

The first stage of the sampling procedure selects establishments while the second stage subsamples ELI's within the chosen establishments. The system is similar for both imports and exports, and unless otherwise stated can be assumed to be identical. For the most part the following discussion will be in terms of imports. The first step in the design is a program which generates the measure of size for an establishment or company. The program is called Tableau I, and the measures of size are called 'max-probs' which are computed as follows. The dollar value on each document is aggregated up to company-TSUSA, company-A number, company-stratum, and company levels. It is also aggregated within a TSUSA, within an A-number, and within a stratum across all companies. A proportion is then calculated for each company-stratum by dividing the aggregated company-stratum dollar value by the aggregated dollar value within that stratum. This "company-stratum prob" is the proportion of dollar value that the company contributes to that particular stratum. A max-prob for each company is the maximum company-stratum prob for that company over all strata. (Since an initial stage of selection is made in exports, a weighted dollar value computed using the weights supplied by Census is aggregated up to appropriate levels.) In addition to a max-prob, a 'max-prob stratum' (the stratum associated with the max-prob) is also assigned to a company. These two items of information are then used in the first stage sample in the following manner. After the companies have been implicitly stratified by max-prob stratum, a systematic PPS selection of companies is made using the max-prob for each company as the measure of size. One reason for using the max-prob is that this results in a company's measure of size being based on the product category for which it is most important. This is appealing since it is product category indexes which are being produced.

When a company is selected in the first stage, it is selected for all its products, including those outside its max-prob stratum. In keeping with the technique presented by Lahiri, the next step should be to use the "relative-prob" (the company-stratum prob divided by the max-prob) as the measure of size and an interval of size 1 when subsampling within the companies. However, this design deviates from the technique in not using an interval of size one in order to incorporate the practical aspects of respondent burden, cost, and publishability requirements. This change still allows each ELI a chance of selection and is accounted for in the weights used in the estimation formula.

Since there is a limit to the number of prices for which a company will respond, it is necessary to specify a maximum number of second stage selections per company. In the export sample design, this number will vary depending on the size of the company and its past history (i.e. the number of quotes already being received in other sections). Only those companies exceeding their maximum will be subsampled. In the import sample design, a more complex algorithm which incorporates the number of consistent ELI's is used. This algorithm is applied to every company, and therefore all companies, regardless of the number of units, are subsampled. It is possible that in using this algorithm the total number of ELI's sent to the field is reduced; however, this cost is more than offset by an increase in the number of consistent ELI's which are sent out for pricing. Table IV contrasts the distributions of the sample ELI's by consistency class in samples produced before the algorithm was introduced and in samples produced after the algorithm was introduced.

As mentioned above, publication requirements also enter into the subsampling. Since the number of ELI's per company is specified, it will usually differ from the expected number of selections in the company. This causes the number of units within a publication category to differ from its expected value. Therefore, to insure that publication requirements are met, all strata with few companies are filled up to specified levels with certainty selections from the "important"

TABLE IV

Effect of Consistency Burden Algorithm
On the Distribution of Sample ELI's
by Consistency Rank

Consistency Rank	Before Consistency Algorithm	After Consistency Algorithm
1	28.3	19.9
2	10.5	10.1
3	17.6	12.2
4	11.4	10.8
5	32.2	47.0
	100.0	100.0

companies in those strata. (The number of probability picks in each of these companies is decreased by one whenever a certainty selection is made).

At this point a single start systematic PPS sample of ELI's is chosen in each company. The relative-prob of each company-stratum is distributed among the ELI's in it in proportion to their dollar value contributions as calculated in Tableau I. This ELI-prob is used as the measure of size in the second stage. The company burden, reduced by the number of certainty selections, is used as the number to select. A listing with the selected companies and their subsampled ELI's is prepared and sent to the field for initiation and quarterly repricing.

Item Sampling

After the sample of exporters or importers and entry level items has been selected a further disaggregation (i.e. partition) is needed to obtain a specific item that can be priced over time. This partition of the entry level item is done in the establishment. The current procedure is a judgmental disaggregation process. The respondent is asked to identify price lines in the ELI and to provide a "representative" item for each price line which can be priced over time. The Bureau has tested a probability disaggregation process in which the ELI is partitioned into subclasses and a PPS selection made among the subclasses using the proportion of trade in the establishment as provided by the respondent as the measure of size. The process continues through successive subdivisions of each selected subclass until an identifiable item that can be priced over time is obtained. Preliminary results indicate that the test was successful; however, the Bureau is conducting further tests to determine whether a broadly or narrowly defined ELI is more suitable to probability disaggregation. Probability disaggregation should eliminate the bias of the judgment disaggregation.

Maintenance and Supplementation

Because of the volatility of the export and import market, the samples have to be replaced more frequently than in other Bureau programs. Current plans are to replace samples every four years. In addition, those sections which have too rapid a decline in responding units to support commodity level indexes until replacement will be supported with supplemental samples. These supplemental samples will be initiated when the number of respondents falls below a specified level needed to support the index.

Estimation

A Laspeyres Index is used to measure the changes in international prices. This type of index measures the change in value of a fixed quantity of commodities over time. The index of price change between time o and time t is given as follows:

$$I_t = \frac{\sum_c P_{ct} Q_{co}}{\sum_c P_{co} Q_{co}} \times 100 = \frac{\sum_c \frac{P_{ct}}{P_{co}} P_{co} Q_{co}}{\sum_c P_{co} Q_{co}} \times 100$$

$$= \frac{\sum_c \frac{P_{ct}}{P_{co}} E_{co}}{\sum_c E_{co}} \times 100$$

where

P_{ct} = price of item c at time t

P_{co} = price of item c at time o

Q_{co} = quantity of item c at time o

$E_{co} = P_{co} Q_{co}$ = value of item c at time o

To produce Laspeyres indexes, the same item is priced over time in a given exporter (importer).

As indicated , the universe of commodities is partitioned into disjoint publishability and residual strata and each publishable subclass is a union of these strata. Therefore, the estimator of the index for any publishable class can be constructed from strata indexes as follows:

$$I_{kt} = \frac{\sum_{h \in k} E_{h,o} I_{ht}}{\sum_{h \in k} E_{h,o}}$$

where

$E_{h,o}$ = total value of exports (imports) for stratum h

I_{ht} = index at time t for stratum h

I_{kt} = index at time t for publishable class k

and the sum is over all strata in class k.

Each stratum level index is a weighted average of price relatives from time o to time t. The basic form of this index is:

$$I_{ht} = \sum_{i}\sum_{j}\sum_{l} W_{ijl}\, r_{ijl\ t\ o} = \sum_{i}\sum_{j}\sum_{l} W_{ijl}\, \frac{P_{ijl\ t}}{P_{ijl\ o}}$$

where

P_{ijlt} = price of item l contained in ELI i in exporter (importer) j at time t

P_{ijlo} = price of item l contained in ELI i in exporter (importer) j at time o

$r_{ijlt\ o} = \frac{P_{ijlt}}{P_{ijlo}}$ = relative price change from time o to time t of item l contained in ELI i in exporter (importer) j

W_{ijl} = weight of item l contained in ELI i in exporter (importer) j.

The weight is a value weight and contains factors for each of the stages of sampling as well as adjustment factors.

To compute indexes for each publication period the following form is developed.

$$I_{ht} = \sum_{i,j,l} W^{o}_{ijl}\, \frac{P_{ijl\ t}}{P_{ijl\ o}} = \sum_{i,j,l} W^{o}_{ijl}\, \frac{P_{ijl\ t-1}}{P_{ijl\ o}}\, \frac{P_{ijl\ t}}{P_{ijl\ t-1}}$$

$$= \sum_{i,j,l} W^{t-1}_{ijl}\, \frac{P_{ijlt}}{P_{ijl\ t-1}}$$

$$= \sum_{i,j,l} W^{t-1}_{ijl}\, \frac{P_{ijl\ t}}{P_{ijl\ t-1}} \times \frac{I_{h\ t-1}}{I_{h\ t-1}}$$

$$= \sum_{i,j,l} \frac{W^{t-1}_{ijl}}{\left(\sum_{i,j,l} W^{t-1}_{ijl}\right)}\, \frac{P_{ijlt}}{P_{ijlt-1}}\, I_{h\ t-1}$$

where

$W^{t-1}_{ijl} = W^{o}_{ijl}\, \frac{P_{ijlt-1}}{P_{ijlo}}$ = updated weight for each item

$W^{o}_{ijl} = W_{ijl}$ = weight of each item at the reference or base period

and $I_{ht-1} = \sum_{i,j,l} W^{o}_{ijl}\, \frac{P_{ijlt-1}}{P_{ijlo}} = \sum_{i,j,l} W^{t-1}_{ijl}$

The index at time t is the weighted average of the price relatives from time period t-1 to t times the index at time t-1 where the weights were normalized over those items for which prices were obtained at time t. The price of the item at time t-1 may be an actual price or an imputed price. Values are imputed for missing prices using the index and the price of the item a time t-1.

Variance estimates will be published along with the indexes for each publishability class. Replication will be used to produce the variance estimates. Half samples will be designated for each stratum and these will be combined to form replicate estimates for each publishable class.

Status

Initiation of commodity classes is scheduled for completion in FY-81. In addition, replacement of samples for commodity areas important to the index is an ongoing process. The current index uses probability samples but produces unweighted indexes. Plans are being implemented to introduce weights and the production of variances into the index. The sampling frame is undergoing investigation to determine if any further modifications to the survey design will yield a higher rate of usable responses. Comparisons of alternative designs to improve efficiency are also underway. Plans are being formulated for assigning identification codes to exporters and including the code on the Shippers Export Declaration. This would allow the export sampling frame to be analyzed in the same way as imports and would further improve the efficiency of the export design, achieving the ultimate goal of a unified export and import sample design.

REFERENCES

[1] The Administration and Operation of the Customs Laws: Customs Procedures with Respect to the Verification of Import Statistics, United States International Trade Commission Publication 820, June 1977.

[2] Supplemental Report on Customs Procedures with Respect to the Verification of Import Statistics: The Accuracy of Import Data, United States International Trade Commission Publication 861, March 1978.

[3] Lahiri, D.B., A Method of Sample Selection Providing Unbiased Ratio Estimates, Bulletin of the International Statistical Institute, Volume XXXIII Part II.