ALLOCATING RESPONDENT BURDEN IN THE INTERNATIONAL PRICE PROGRAM

Gwyn R. Ferguson and Marvin Kasper, Bureau of Labor Statistics

Introduction

The purpose of the International Price Program (IPP) is to produce indexes which measure price change for virtually all goods which are imported into or exported from the United States.[1] To publish these indexes, the IPP collects prices for imported and exported products. It would not be practical for the IPP to collect data from every firm which imports and/or exports goods because the resources and associated cost factors would be prohibitive. Thus the IPP has implemented survey sampling techniques to select a group of specific items which can be repriced over time to provide the price data for publishing the IPP indexes.[2] This group of specific items is selected using a three-stage sample design. The first stage is the selection of companies. The second stage is the selection of product areas traded by each company. The third stage is the selection of specific items within each product area. Because of the need to expand program coverage and the detailed levels at which IPP indexes are produced, it is not practical for the IPP to select items for repricing across all product areas at one time. Instead, the IPP selects independent samples from several broad product areas. Usually only one or two samples are selected each year.

Many of the sampled companies import and/or export a variety of items in a large number of product areas. The number of products for which a sampled company is asked to report prices is called the reporting burden of the company. The reporting burden of companies trading in a variety of commodities must be carefully monitored to maximize the usefulness of the collected data in the production of indexes and to enhance the likelihood of cooperation and participation of these companies.

Since samples are generally selected for trade in a broad product area, there is a problem in deciding what the reporting burden for each company should be in each area so that the reporting burden across all product areas is not too large. If a company's total burden is too large, that company may decide that it doesn't have the resources to provide any or all of the requested prices. However, if the total burden is reduced, the company might be willing to participate in the survey. Therefore, we developed an algorithm for determining the total burden of a company and for allocating this burden across product areas. The study to develop this algorithm concentrated on importers because the information necessary to develop this burden across product areas (the total reporting burden) is not available for exporting companies.

In order to determine the algorithm for setting a maximum burden for some major importers, an analysis of the 500 largest importers by dollar value of imports was conducted. Based on this analysis, an algorithm was developed for setting a maximum burden across all product areas and distributing the total burden across broad product groupings. This algorithm was used to set burdens for the companies which were studied.

Source Of Data

The data used for the analysis were obtained from the U.S. Customs Service. The 500 largest importers (11-digit consignee identifiers) by total dollar value for the period October, 1978 through July, 1979 were identified. The following information was obtained for each import transaction conducted by this group of importers:

- consignee identifier,
- consignee name and address,
- transaction c.i.f. dollar value,
- Tariff Schedule of the United States Annotated (TSUSA) [3] under which the transaction was filed, and
- the month in which the transaction occurred.

The Schedule A number associated with the reported TSUSA was added to this information to complete the data elements needed for the analysis. The Schedule A classification is a nested classification of products in which 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. Each 7-digit Schedule A number corresponds to one or more TSUSA numbers. Since the product area from which a sample is drawn generally consists of several one or two digit Schedule A categories, the study was restricted to looking at imports within each of the one digit and two digit product categories.

Analysis of the Data

The data were examined based on four factors, diversity of imports, distribution of trade, consistency of trade, and coverage of the universe. The first factor was the diversity of imports or the number of product areas traded by these companies. If these companies trade in a variety of product areas, there is a need to control the total reporting burden for these companies by limiting the number of quotes, or products, for which data is requested as each product area is sampled.

A frequency distribution of the number of one digit sections of the Schedule A imported by each of these major importers is shown in Table A. 54% of these companies import products in six or more one digit areas while only 19% import in only one or two one digit Schedule A groupings. These companies import products in an average of 5 one digit areas and 17 two digit areas. This indicates that imports by these companies are rather diverse.

The second factor which was examined was the distribution of trade for each of these importers. Companies are selected for participation in the International Price Program based on the product area in which the company was most important. This factor was examined to determine if the area which was most important to the company was the same as the area in which the company was most important to the IPP. If so, the method for allocating the total burden could use the percent of company imports in the various product classifications while retaining quotes in areas needed for publication of the IPP indexes.

Table B shows the distribution of the percentage of company dollar value in the product area in which the company was most important. For example, suppose company A was more important in Section 5 than in any other section and that 45% of its imports are in Section 5. Then company A would be included in the 40.0-49.9 percent range. As we expected, 87.4% of the companies had more than 70% of their imports in the product area in which they were most important. However, there were 7 companies in which less than 5% of the company's imports were in the section of largest importance. Even though the percent of the dollar value of imports might have been low in the area for which the company was important, the burden algorithm would need to ensure that the low volume area received an allocation of one or more quotes.

The consistency of trade was the third area of interest in the analysis. Data from past samples have shown that the more frequently an item was imported the the more likely that a price will be obtained from the company for the item. A consistency rank was assigned to each company-TSUSA based on the number of months and the number of quarters in which the owner imported in the TSUSA. A company-Schedule A was assigned the maximum rank of any company-TSUSA contained in it. A company was assigned the highest consistency rank of any of its TSUSA's. A consistency rank of 6 represents the most frequent trade while a rank of 1 indicates less frequent trade. A company-Schedule A or company is defined to be "consistent" if it is assigned a consistency rank of 3 or more. Otherwise, the company-Schedule A or company is defined to be "inconsistent".

Table C shows the distribution of companies by consistency rank. 92.4% of the companies imported at least one product of rank 6 while only 0.4% (2 companies) had no imports above rank 2. One of these inconsistent companies only traded one product while the other one traded infrequently in several product areas. Table D shows the distribution of company-Schedule A's by consistency rank. 52.5% of the company-Schedule A's had a consistency rank of 1 but they only represented 1.6% of the import dollar value of these 500 importers. 17% of the company-Schedule A's had a consistency rank of 6 and they accounted for 87.5% of the total dollar value. Past studies have shown that the lower rank company-Schedule A's are less likely to yield a price, as well as the fact that they are usually less important to the IPP indexes. Thus, the burden algorithm needs to consider the consistency rank of a company's trade of products in each one digit section when allocating the burden.

The fourth area of analysis was the coverage of import trade. In order to determine the magnitude of imports by these major companies, the total dollar value of trade of these companies was compared to figures published by the Bureau of the Census for all imports [4]. This group of major importers represented 58.3% of the total dollar value of trade for the period October, 1978 through July, 1979 (Table E). These large companies represented over 97% of imports in mineral fuels (i.e. oil, natural gas, and coal) and 64% of imports in machinery and transport equipment. This was an expected result since imports in oil and automobiles are concentrated in a few very large companies. A cumulative frequency distribution of these large importers showed that the largest 11 companies represented more than 10% of the total trade while 140 companies were required to raise the total percentage of trade from 40% to 50% (Table F). These tables showed that the majority of the dollar value of imported goods was imported by a relatively small group of companies.

Burden Algorithm

Based on the analysis of the largest 500 importers, an algorithm was developed to determine the reporting burden for a company. The algorithm is composed of two parts. The first part is a technique for determining the maximum total reporting burden for a company and the second part is a technique for allocating the reporting burden to different classes of commodities (in this case different one digit groupings of the Schedule A). Several factors which are considered in the reporting algorithm are as follows:

1. Importance of the company to specific publishability classes (two digit Schedule A groupings).
2. Frequency with which a company trades in specific commodities.
3. Diversity of products traded by a company.
4. Importance of a two digit Schedule A grouping to total import trade.
5. Importance of the company in foreign trade.

By considering the importance of a company to publishability classes and the frequency with which a company trades commodities, the number of collected usable prices should increase. In addition, by considering the diversity of products traded by a company, the desired quotes are spread over the entirety of the company's product areas. By allowing for the importance of the two digit grouping, the algorithm incorporates program objectives of publishing indexes for important groupings of the Schedule A. By considering the importance of the company, quotes are allocated based on the importance of the company to the IPP indexes.

The maximum total burden for a company is determined by the total dollar value of all products imported by the company. It is defined as follows.

$$MB = \begin{cases} 30 & \text{if } \$1{,}000{,}000{,}000 \leq T \\ 25 & \text{if } \$\ 500{,}000{,}000 \leq T < \$1{,}000{,}000{,}000 \\ 23 & \text{if } \$\ 250{,}000{,}000 \leq T < \$\ 500{,}000{,}000 \\ 19 & \text{if } \$\ 100{,}000{,}000 \leq T < \$\ 250{,}000{,}000 \\ 15 & \text{if } T < \$\ 100{,}000{,}000 \end{cases}$$

where MB = the maximum total burden for the company, and
T = the total dollar value of imports for the company.

The total dollar value of the company was used to determine the maximum burden because companies with larger dollar values are more likely to be selected in the IPP samples. It is also desir-

able to distribute the weight of these companies in the IPP indexes across several products.

Once the maximum burden was determined for a company, an allocation for each one digit section of the Schedule A was calculated. The first part of this allocation is based on the number of products and two digit Schedule A groupings traded by a company. This formula also considered the consistency of trade of products within the Schedule A numbers and within the two digit Schedule A groupings. The consistent company-Schedule A's and two digit Schedule A groups were weighted more heavily than the inconsistent ones since the past response rates have been more favorable for products in the consistent categories. Since past experience indicated that the number of Schedule A numbers traded by a company was more important to the overall burden than the number of two digit Schedule A areas traded, the actual counts of Schedule A numbers were weighted more heavily in the formula.

$$B_j = \text{CEIL}\,[\,0.75 * (NCE_j + 0.2 * NIE_j) + 0.25 * (NCS_j + 0.2 * NIS_j)]$$

where B_j = preliminary burden for one digit section j of the Schedule A for this company,

CEIL(x) = the smallest integer greater than or equal to x,

NCE_j = the number of consistent company-Schedule A's traded by this company in one digit section j,

NIE_j = the number of inconsistent company-Schedule A's traded by this company in one digit section j,

NCS_j = the number of two digit sections of the Schedule A traded in which the company had at least one consistent company-Schedule A in one digit section j,

and NIS_j = the number of two digit sections of the Schedule A traded in which the company had no consistent company-Schedule A's in one digit section j.

The allocations for each one digit section are then summed across the one digit sections to obtain the total preliminary burden for the company, TB.

$$TB = \sum_{j=0}^{9} B_j$$

If TB < MB, then the final burden for the company was TB and the burden for each one digit section of the Schedule A was B_j. However, if TB > MB, then the final burden for the company was MB and this burden was allocated to the various one digit Schedule A sections according to the following algorithm.

A measure reflecting the importance of the company within one digit section j of the Schedule A, R_j, was calculated for each one digit section in which the company traded. This measure reflected the relative importance of the company and Schedule A's traded in the one digit category j to other items in different one digit categories. The R_j were used to allocate the final burden for the company across the one digit Schedule A groupings.

$$R_j = \sum_{k \varepsilon j} I_k \sum_{e \varepsilon k} P_{e,k} * f_{k,c}$$

where I_k = the proportion of the dollar value of two digit section k to all U.S. imports.

$P_{e,k}$ = the proportion of the dollar value of the company-Schedule A e in two digit section k,

and $f_{k,c}$ = the probability of a good response for a company-Schedule A number in two digit section k with consistency rank c,

The final allocation for each section of the Schedule A is determined by proportionally allocating the maximum total burden across the one digit sections by the R_j. There is a constraint that the burden for any one section j as determined by this algorithm must be less than or equal to B_j.

Results

Each of the 500 major importers was assigned an overall reporting burden using the burden algorithm. Imports in automobiles and mineral fuels tend to be concentrated within a few very large importers and currently do not pose as much of a respondent burden problem for the IPP as other areas of trade. Thus imports in these product areas were not considered in assigning and allocating the respondent burden. A distribution of the total burden for these companies is given in Table G. The average company among these 500 was assigned 16 quotes across all product areas. The largest burden assigned using the algorithm was 35 total quotes across all the one digit Schedule A sections. 71% of the companies received a total reporting burden between 5 and 24 quotes. Table H shows the distribution of the assigned reporting burden across each one digit section of the Schedule A. Over 52% of the assigned quotes were allocated to Sections 6 and 7 (manufactured goods classified by material and machinery and transport equipment other than automobiles). These two sections represented 50.4% of the total import dollar value for October, 1978 through July, 1979.

Conclusion

This algorithm controls the total reporting burden of a company while allocating this burden in a way that will increase the usefulness of the data collected to index production. By considering the importance of a company to two digit

Schedule A groups and the frequency with which a company trades commodities, the number of collected usable prices should increase. In addition, by considering the diversity of products traded by a company, the desired quotes can be spread over the entirety of the company's product areas. By allowing for the importance of two digit Schedule A groups, the algorithm can incorporate program objectives. While not the only possibility, this algorithm does seem to offer a rational approach to large company burden allocation. In the future, this algorithm will continue to be refined to better meet the needs of the IPP. Current plans also call for applying a reporting burden algorithm to a larger group of importing companies and to a group of the largest exporting companies.

References

[1] Handbook of Methods, United States Department of Labor, Bureau of Labor Statistics Bulletin 2134-1, 1982

[2] Kasper, M. and Pratt, R.J., Surveying International Prices, 1978 Survey Research Methods Section Proceedings of the American Statistical Association, pp. 499 - 504.

[3] Tariff Schedules of the United States Annotated, United States International Trade Commission Publication 843, 1977.

[4] U.S. General Imports and Imports for Consumption, U.S. Department of Commerce, Bureau of the Census, FT135, December 1982.

TABLE A

Distribution of One Digit Schedule A Numbers Imported

Number of One Digits Imported	Percent of Companies
1	9.6
2	9.6
3	5.6
4	9.8
5	11.8
6	15.4
7	15.4
8	13.4
9	7.8
10	1.6
Total	100.0

TABLE B

Distribution of Companies over Percent of Imports in Product Area in which the Company is Most Important

% of Company Dollar Value in Product Area where Company is most Important	Percent of Companies
0.0 - 4.9	1.4
5.0 - 9.9	1.0
10.0 - 19.9	2.8
20.0 - 29.9	1.0
30.0 - 39.9	0.6
40.0 - 49.9	1.8
50.0 - 59.9	2.0
60.0 - 69.9	2.0
70.0 - 79.9	4.0
80.0 - 89.9	8.0
90.0 - 100.0	75.4
Total	100.0

TABLE C

Distribution of Companies by Consistency Rank

Consistency Rank	Percent of Companies	Percent of Dollar Value
1	0.0	0.0
2	0.4	0.1
3	1.4	1.2
4	1.2	0.4
5	4.6	2.1
6	92.4	96.1

TABLE D

Distribution of Company-Schedule A's by Consistency Rank

Consistency Rank	Percent of Company-Schedule A's	Percent of Dollar Value
1	52.5	1.6
2	11.3	1.0
3	5.8	2.5
4	5.3	1.1
5	8.0	6.3
6	17.1	87.5

TABLE E

Percent of Total Universe Dollar Value
Represented by the Largest 500 Importers

Schedule A One Digit Section	Percentage
0 - Food and live animals	43.7
1 - Beverages and tobacco	23.6
2 - Crude materials, inedible, except fuels	35.4
3 - Mineral fuels, lubricants and related materials	97.8
4 - Oils and fats, animal and vegetable	40.2
5 - Chemicals and related products, NSPF	35.0
6 - Manufactured goods classified chiefly by material	33.2
7 - Machinery and transport equipment	64.6
8 - Miscellaneous manufactured articles, NSPF	23.3
9 - Commodities and transactions not classified elsewhere	29.4
Total	58.3

TABLE F

Cumulative Frequency Distribution
of the Largest 500 Importers by Dollar Value

Percent of Total Import Dollar Value	Number of Companies
10	11
20	29
30	68
40	136
50	276
58.3	500

TABLE G

Distribution of Assigned Reporting Burden

Total Reporting Burden	Percent of Companies
0 - 4	14
5 - 9	13
10 - 14	12
15 - 19	19
20 - 24	27
25 - 29	11
30 - 35	4
Total	100

TABLE H

Distribution of Assigned Reporting Burden
by One Digit Schedule A Section

Schedule A Grouping	Percentage
0 - Food and live animals	7.7
1 - Beverages and tobacco	2.8
2 - Crude materials, inedible, except fuels	6.3
4 - Oils and fats, animal and vegetable	0.6
5 - Chemicals and related products, NSPF	9.6
6 - Manufactured goods classified chiefly by material	21.3
7 - Machinery and transport equipment	31.0
8 - Miscellaneous manufactured articles, NSPF	15.2
9 - Commodities and transactions not classified elsewhere	5.5
Total	100.0