

EXPORT STRATIFICATION ALTERNATIVES IN THE INTERNATIONAL PRICE PROGRAM

Karen K. Hess and Gwyn R. Ferguson, 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 Entry Level Items (ELI's-product areas traded by each company). The third stage is the selection of specific items in the ELI.

Because of imperfections of the sampling frame, out of scope units are selected. For example, exporters frequently discontinue or change their product categories. The item selected is then not eligible for repricing and is therefore out of scope. Other reasons for a sampling unit being out of scope include unobtainable market prices, lack of periodicity, and misclassification of the commodity.[3]

To improve the efficiency of the survey and decrease the percentage of sampled exporting companies which are out of scope of the survey, the feasibility of implementing a stratified sample design for exports was investigated. The strata would be defined by characteristics of the frame that are associated with exporters that are in scope of the survey. In scope companies are defined as having sampled products eligible for the IPP indexes.

Export Frame

Exported goods are classified using two seven digit classification systems, the Schedule B and the Schedule E. The Schedule E is a nested system based on the Standard International Trade Classification (SITC). Each seven digit Schedule E category is mapped in a one-to-one relationship to a seven digit Schedule B category. The classification is also used to define the Entry Level Item (ELI). An ELI contains a homogeneous group of one or more Schedule B categories. The ELI's are then aggregated to the SITC section or set of subclasses for which indexes are to be published.

Exporters must report their transactions to the U.S. Customs Service on Shipper's Export Declarations (SED's). When an exporter files a Shipper's Exporter Declaration, he classifies the item exported using the Schedule B classification. Each SED contains information about the name and address of the exporter, the Schedule B number under which the exporter classified the goods being exported, the f.a.s. U.S. port valuation of the goods, and the month and year of

exportation. The Customs Service forwards the SED's to the Bureau of the Census. The Bureau of the Census selects a sample of the documents filed during a specified time period for specified sections of the SITC to form the IPP sampling frame of exports. Usually only one sample is selected each year from a select group of product areas. The samples are given a numeric name based on chronological order. For example, the eighth exporter sample was drawn from calendar year 1980 data and covered SITC section 0 (food and live animals), parts of sections 2 (crude materials), and parts of section 5 (chemicals), while the ninth exporter sample was selected from the first nine months of calendar year 1981 data and included several other SITC sections.

The Bureau of Labor Statistics (BLS) then groups the actual transactions selected into establishments according to the similarity of the names and addresses on the documents. The ELI number corresponding to the reported Schedule B number is added to the information from the Census Bureau to complete the information needed for sampling. The sampling frame is then aggregated both within exporters and across exporters to the Schedule B number level, the ELI level, and the exporter level.

Stratification Variables

We identified three types of stratification variables to be studied. They were the monthly distribution of trade, the frequency of reporting trade, and the value of trade. For the monthly distribution of trade we defined three variables: number of months traded, number of quarters traded, and a consistency rank currently used for stratifying the import frame which combines the number of months traded and the number of quarters traded. The number of documents reported was the only variable defined for the frequency of reporting trade. The two variables we defined in the area of value of trade were the actual dollar value of a Schedule B within an exporter (exporter/Schedule B) and the exporter/Schedule B prob. The exporter/Schedule B dollar value is the unweighted total dollar value of all products in our sampling frame within the Schedule B classification which were exported by the company. The exporter/Schedule B prob is defined as the exporter/Schedule B dollar value divided by the universe dollar value for the aggregated publishability set corresponding to that Schedule B.

For each variable, 4 - 12 alphabetic ranks were defined. A rank of A for the monthly distribution of trade implies infrequent trading while the ranks of B through L imply more frequent trade. Similarly, for the number of documents a rank of A is one document and a rank of K is more than fifty documents. The ranks assigned to the criteria of dollar value and exporter/Schedule B prob refer to a range of these values. Therefore, in this study the term "higher rank" refers to both an increasing alphabetic order and an increasing frequency of trade

or value. Complete definitions of the six variables are given in Table A.

For each stratification criteria a rank was first determined for each exporter/Schedule B in the frame. The maximum rank of the exporter/Schedule B's within an exporter/ELI was then assigned to the exporter/ELI. Using the same procedure, a rank was assigned to each exporter for each criteria corresponding to the maximum rank of any Schedule B number within the company for that criteria.

Analysis of the Stratification Alternatives

In order to analyze the possible effect of stratification on the percentage of out of scope units in a sample, the distribution of in scope and out of scope units by rank was compiled using the seventh, eighth, and ninth exporter samples. The seventh and eighth exporter samples were each selected from twelve months of data while the ninth exporter was sampled using only nine months of data. The samples also differ by sectors of coverage. However, by looking at each sample individually it was clear that the in scope rate distribution was similar for each variable regardless of the length of the sampling period or SITC section sampled. Therefore, these differences do not preclude comparison of the in scope rate distributions. The summary tables are thus given for the total of the seventh, eighth, and ninth exporter samples.

Tables B and C show the total and percentage of exporters and exporter/ELI numbers which were in scope and which were out of scope for the variable of number of months traded. These tallies are given for the total of the three samples studied. The same kinds of tables were produced for each of the other variables. Table B shows that, at the exporter level, in scope rates improve as the ranks increase. The improvement is most obvious between the lowest rank and the group of remaining ranks. As seen, rank A has an in scope rate of 42.6% while rank B has an in scope rate of 62.1%, an increase of 19.5% while the change between the second rank (B) and the highest rank (L) is only 15.0%. To further illustrate this effect, the out of scope rate across the number of months was plotted as shown in Table D. A quadratic curve was then fitted to the points. Again, the change is greater between the lowest rank of one month and the remaining ranks of two through twelve months.

At the exporter/ELI level there also is a significant change between rank A and rank B. As illustrated by Table C, the in scope rate change from rank A to rank B is 18%. The number of months variables provided a very similar in scope rate pattern as the number of quarters and consistency rank variables. Also, for all three monthly distribution of trade variables, there was a significant difference between the in scope rate for rank A and the in scope rate for rank B while the rate of increase between each of the higher ranks was less. Thus we decided to merge the ranks for each of these variables into two groups. Group 1 for each variable consisted of rank A while group 2 consisted of all other ranks. Since the consistency rank A consists only of units traded during one month in one quarter, this grouping caused the consistency

variable to correspond to the number of months variable. We then compared the in scope rates for the number of months traded to those for the number of quarters traded for these two groups. From this comparison, we noted that there is not as much of an increase in the in scope rate between group 1 and group 2 for the number of quarters variable. Since the number of months variable allows us to include those units which trade in 2 or 3 months during one quarter among the group 2 units, we chose the number of months alternative as the preferred stratification variable among the monthly distribution of trade variables.

The in scope distributions for the number of months traded and the number of documents reported were very similar. A drawback to the number of documents, i.e. frequency of reporting trade variable, is the method of recording companies that report export trade data to the Customs Service by computer tape instead of actual paper copies of the SED's. It is difficult to determine exactly how many "paper documents" would have been filed by the company for a particular Schedule B number and attempts to try to capture this information would require extensive modifications to our computer system.

The initial analysis of the value of trade variables showed less correlation for these variables and the in scope response than was shown for the number of months traded. Thus we excluded these variables from further consideration as stratification alternatives.

Based on this analysis, the preferred variable for stratification was the number of months. We decided to form two stratification groups: Group 1 - one month of trade and Group 2 - more than one month of trade. However, before implementing a stratification technique into our sample design we needed to see what gain, if any, can be achieved by using stratification. In particular, we needed to determine if our current sample design is already selecting all the available units in Group 2.

Feasibility of Proposed Stratification Criteria

Our current sample design incorporates two stages of sampling prior to initiation of the exporter. In the first stage, exporters are selected to participate in the program. During the second stage, specific ELI's are chosen within each exporter. A third stage of sampling, selection of a specific product within the exporter/ELI, is conducted during the initiation visit to the exporter.

Table E shows the distribution of companies in each stratification group for the number of months variable for the combined seventh, eighth, and ninth exporter samples. This table shows the number and percent of exporters within each group in the original sampling frame and after selection of the first stage sample. From Table E we see that our current sampling technique is selecting an average of 51.5% of the Group 1 companies and 48.5% of the Group 2 (more than one month of trade) companies for the sample. However, an average of only 67.7% of the Group 2 companies available from the frame are being selected.

Now consider the maximum potential gain from stratifying at the first stage of sampling. Based on the data from Table B the average in scope rate for the 3 samples is 43% for Group 1 and was computed to be 70% for Group 2. Since this table does not include any units which are still pending collection, we calculated the expected number of in scope units for the three original samples assuming the rates would be the same for the pending units. We then determined the number of companies in each sample if we selected all Group 2 companies in the frame with certainty and used the Group 1 companies to complete the overall sample size for the original samples. Based on the number of companies in each group, we calculated the expected number of in scope companies for each of these stratified samples. This is summarized in Table F.

The average number of in scope ELI's per in scope company is 1.7 so the number of expected in scope company/ELI's for the original sample is 4246 (2498 * 1.7) and the number of expected in scope company/ELI's for the stratified sample is 4719 (2776 * 1.7). The expected increase of in scope company/ELI's is computed as follows:

Expected increase of in scope company/ELI's

$$= \frac{\text{expected \# of stratified company/ELI's} - \text{expected \# of original company/ELI's}}{\text{expected \# of original company/ELI's}}$$

$$= 4719 - 4246$$

$$= 473.$$

Therefore the expected gain due to stratification at the first stage of sampling is

$$= \frac{\text{\# of additional in scope company/ELI's}}{\text{\# of in scope company/ELI's}} * 100$$

$$= \frac{473}{4246} * 100$$

$$= 11.1\%$$

Thus, by stratifying the export frame at the first stage of sampling, we can expect to increase the number of in scope company/ELI's by 11.1%.

An analysis of the second stage of sampling showed that the stratification of exporter/ELI's into the two groups resulted in a maximum expected gain of 1.4%. This small gain of in scope units does not warrant stratification at the second stage of sampling.

Conclusion

Stratifying the export frame before selecting companies would have produced a significant increase in the expected number of in scope company/ELI's for these three previous samples. Thus we decided to stratify future samples at the first stage of sampling into two groups. The first group would consist of companies in which every Schedule B number was traded during only one month. The second group would consist of companies in which at least one Schedule B number

was traded during more than one month. By using this technique we expect to increase the number of useable products available for index production by approximately 11%.

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] Carpener, J. Finley. "Error Analysis in the International Price Program." Paper presented at the American Statistical Quality Control Technical Conference, Chicago, 1978.

TABLE A

Definitions of the Alternative Criteria

I. Number of Months Traded

<u># of Months</u>	<u>Rank</u>
1	A
2	B
3	C
4	D
5	E
6	F
7	G
8	H
9	I
10	J
11	K
12	L

II. Number of Quarters Traded

<u># of Quarters</u>	<u>Rank</u>
1	A
2	B
3	C
4	D

III. Importer Consistency Rank

	Total Quarters Traded				Rank
	1	2	3	4	
					A
1	1				B
2	2	4			C
3	3	6	9		D
4		8	12	16	
5		10	15	20	E
6		12	18	24	
7			21	28	
8			24	32	
9			27	36	
10				40	F
11				44	
12				48	

IV. Number of Documents Traded

# of Documents	Rank
1	A
2	B
3	C
4	D
5	E
6	F
7	G
8	H
9	I
10 - 49	J
50 or more	K

V. Exporter/Schedule B Prob Ranges: The exporter/Schedule B prob is defined as the exporter/Schedule B dollar value divided by the universe dollar value for the aggregated publishability set corresponding to that exporter/Schedule B.

Prob Range	Rank
0.000000 - 0.000009	A
0.000010 - 0.000050	B
0.000051 - 0.000100	C
0.000101 - 0.000500	D
0.000501 - 0.001000	E
0.001001 - 0.005000	F
0.005001 - 0.010000	G
0.010001 - 0.100000	H
0.100001 - 1.000000	I

VI. Exporter/Schedule B Dollar Value Ranges

Dollar Value Range	Rank
0 - 500	A
501 - 1,000	B
1,001 - 2,000	C
2,001 - 3,000	D
3,001 - 5,000	E
5,001 - 7,500	F
7,501 - 10,000	G
10,001 - 25,000	H
25,001 - 50,000	I
more than 50,000	J

Table B

Distribution of Companies by Number of Months Criteria

Rank	In Scope		Out of Scope		Total	
	#	% of C	#	% of C	#	%
A	879	42.6	1,186	57.4	2,065	53.7
B	399	62.1	244	37.9	643	16.7
C	241	71.1	98	28.9	339	8.8
D	171	71.2	69	28.8	240	6.2
E	135	79.4	35	20.6	170	4.4
F	63	70.0	27	30.0	90	2.3
G	65	78.3	18	21.7	83	2.2
H	48	81.4	11	18.6	59	1.5
I	46	73.0	17	27.0	63	1.6
J	18	85.7	3	14.3	21	0.6
K	32	82.1	7	17.9	39	1.0
L	27	77.1	8	22.9	35	0.9
Total	2,124	55.2	1,723	44.8	3,847	100.0

Table C

Distribution of Company/ELI's by Number of Months Criteria

Rank	In Scope		Out of Scope		Total	
	#	% of C	#	% of C	#	%
A	1,868	37.8	3,079	62.2	4,947	63.6
B	643	55.8	509	44.2	1,152	14.8
C	364	65.4	193	34.6	557	7.2
D	225	64.5	124	35.5	349	4.5
E	159	69.4	70	30.6	229	2.9
F	75	59.5	51	46.5	126	1.6
G	68	66.7	34	33.3	102	1.3
H	59	67.8	28	32.2	87	1.1
I	61	67.8	29	32.2	90	1.2
J	27	84.4	5	15.6	32	.4
K	35	68.6	16	31.4	51	.7
L	35	67.3	17	32.7	52	.7
Total	3,619	46.6	4,155	53.4	7,774	100.0

Table D

Number of Months Traded
Versus
Company Out of Scope Rate

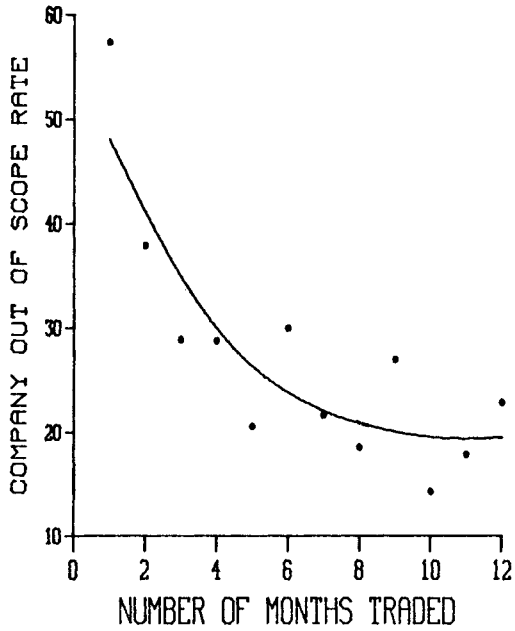


Table E

Distribution of Companies
by Stratification Groups

Group	Frame		Sample		% of Frame
	#	%	#	%	
1	8,569	72.9	2,293	51.5	26.8
2	3,189	27.1	2,159	48.5	67.7
Total	11,758	100.0	4,452	100.0	37.9

Table F

Distribution of Companies
by Stratification Groups

Group	Original Samples		Stratified Samples	
	#	Expected # of in scope	#	Expected # of in scope
1	2,293	986	1,263	543
2	2,159	1,512	3,189	2,233
Total	4,452	2,498	4,452	2,776