ALTERNATIVE SAMPLE DESIGNS IN THE INTERNATIONAL PRICE PROGRAM

Richard J. Pratt and Gwyn R. Ferguson, Bureau of Labor Statistics

INTRODUCTION

The International Price Program (IPP) produces an export price index and an import price index.[1] When the sampling methodology for the IPP was first developed only export data were available. In addition, because there was no exporter identifier on the Shippers Export Declaration (SED) only a sample was drawn from the Bureau of the Census files. As a result, the sample design that was proposed for the IPP was tailored to a sample of data. When import data became available, a similar method was introduced even though an import identifier is included on each import document and therefore the entire file of documents is used as the IPP sampling frame. It now appears that an exporter identifier will be added to the SED and the entire file of export data will be made available to us. As a result we are comparing the current techniques to some possible alternatives with the goal being a unified system designed to yield the best results for both imports and exports.

The four designs which were compared were:

1) Design 1 - our present design - a modification of a design first presented by Lahiri [2],

2) Design 2 - the theoretical design presented by Lahiri,

3) Design 3 - independent selection of products within strata,

and 4) Design 4 - selection of companies using independent selection of products.

SCOPE

All the terms in this discussion will be related to import data since only import data were tested. There are analagous items in exports though the terms may differ.

The following description of the import product classification system should help explain the program requirements. Imports are classified in two seven digit systems, the Tariff Schedule of the United States Annotated (TSUSA)[3] and the Schedule A.[4] The Schedule A is a nested system 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 A are further partitioned using the more detailed seven digit categories of the TSUSA. When an importer files a document, he classifies the item imported using the TSUSA classification.

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 Schedule A level to an overall one digit level. Those classes containing no publishable subclasses are defined to be publishability strata. The items in a publishable class not contained in any publishability stratum are formed into strata called the residual strata of the given class. The publishability and residual strata partition the commodities into disjoint sets. They are the strata used to construct the survey design.

The classification system is also used to define entry level items (ELI's). These are commmodity classes with which a field representative enters an establishment. He then proceeds to obtain a specific item in each class which is priced over time. An ELI contains homogeneous items with respect to price change. The seven digit A-number is the ELI.

In the IPP we are trying to price items over time and therefore are very concerned about a company's consistency of trade. The most serious problem that our survey design must address is the volatility in the international trade market and as a result a company's consistency of trade in a commodity was considered very important in our analysis.

FRAME

The data used in our program is obtained from the U.S. Customs Bureau for imports and from the Bureau of the Census for exports. With the addition of an exporter identifier to the SED beginning in 1980, both the export and import sampling frame will contain all documents for the reference period in the sections of trade requested. The information on the import document file includes the company identification number, name and address, the transaction dollar value, the TSUSA under which it is filed, and the month in which the import occurred. We attach an A-number and stratum code to complete the information necessary for sampling.

The analysis of the various sample designs was conducted using two sets of import data. These were the files used to select the two most current importer samples. The first file is the set of all transactions in products classified in Section 7 of the Schedule A for the period January through June, 1978. The Fifth Importer Sample was drawn from this data. Calendar year 1978 data for transactions in products classified in Section 8, which was used to select the Sixth Importer Sample, comprises the second set of data. Section 7 is machinery and transport equipment while miscellaneous manufactured articles are classified in Section 8. These two sections are two of the largest Schedule A sections both by dollar value and number of transactions. For the six month reference period, Section 7 contained almost one million documents and for the full year of 1978, Section 8 contained approximately 1.86 million transactions. Previous sets of data have contained approximately 1.5 million documents for a full year's data. These sets of data are similar to each other and to other data received from Customs in that the majority of products imported by a company is traded infrequently and represents a small percentage of the dollar value imported in the section. The majority of the dollar value in each file is represented by a small percentage of the company-products traded. The major difference between the two sets of data is the length of the reference period.

FACTORS OF ANALYSIS

The analysis is based on four factors important to the IPP. The first factor is the number of companies in the sample. This is important because there are cost restraints on the program. The expense involved in initiating a company is one of the major costs of the program and therefore the number of companies initiated must be controlled.

The second factor in the analysis is the number of company-ELI's in the sample. Since the purpose of the program is to produce indexes for product sections it is necessary to attain minimum publishability requirements in each stratum. The greater the coverage and the number of company-ELI's in the sample, the better the chance of receiving enough prices to publish indexes by section.

The third factor in the analysis concerns expected and achieved response rates. Data from previous samples has shown that the more frequently an item is imported the better its response rate. We take advantage of this by assigning a consistency rank to each company-TSUSA, company-ELI, company-stratum, and company in the following manner. A score is assigned to a company-TSUSA based on the number of months and number of quarters during the reference period in which the owner imported in a given TSUSA. This score is converted to a consistency rank for that owner-TSUSA. After consistency ranks have been assigned for all owner-TSUSA's, consistency ranks are assigned at the owner-A-number, owner-stratum, and owner levels by assigning the maximum rank over all ranks in the appropriate sublevel. The frequency distribution of the company-ELI's by consistency rank is the third factor of analysis.

The fourth factor important to the analysis is the distribution of the number of company-ELI's per company or reporting burden. Many importers trade in a wide variety of product areas. If prices were requested for most of their products, a company might decide it didn't have the resources to provide any or all of the information. However, if only a few prices are requested, the company might be more willing to cooperate.

PRESENT DESIGN - DESIGN 1

The present design is based on a technique first presented by Lahiri. Several modifications have been incorporated to increase the number of consistent ELI's sampled and control the company burden. The design consists of two stages of sampling. In the first stage, companies to be visited are selected. ELI's within each chosen establishment or company are selected in the second stage.[5]

The first step in the design is to calculate for each establishment a measure of size, called a "max-prob". It is computed as follows. The dollar value on each document is aggregated to company-TSUSA, company-A-number, company-stratum, and company levels. It is also aggregated to TSUSA, to A-number, and to stratum levels across all companies. The ratio of the company-stratum dollar value to the aggregated stratum dollar value is calculated for each company-stratum. This ratio is called the "company-stratum prob" and is the proportion of dollar value that the company contributes to the particular stratum. A max-prob for each company is defined as the maximum company-stratum prob for that company across all strata. A "max-prob stratum" is also assigned to each company. It is the stratum associated with the max-prob. After implicitly stratifying the companies by max-prob stratum, the first stage sample is drawn by making a systematic probability proportionate to size (PPS) selection of companies using the max-prob for each company as the measure of size.

When a company is selected in the first stage, it is selected for all ELI's, including those outside its max-prob stratum. In order to control each company's burden, the ELI's within every company are sampled. The number of ELI's to select for each importer is calculated by tallying the total number of ELI's and the number of consistent ELI's. A company-ELI is defined as consistent if it has a consistency rank greater than or equal to three. Then the burden is determined by

$$B = \begin{cases} 1 & T = 1 \\ \text{Min}\left\{\text{Max}\left[\text{Consist} + \text{Ceil}\left(\frac{T - \text{Consist}}{2}\right), 2\right], \text{RF}\right\} & T \neq 1 \end{cases}$$

where T = total number of ELI's in the company,

Consist = total number of consistent ELI's in the company,

Ceil(x) = smallest integer greater than or equal to x

and

$$\text{RF} = \begin{cases} 5 & T < 11 \\ 6 & 11 \leq T < 25 \\ 7 & 25 \leq T \end{cases}$$

This algorithm is based on the observation that the more consistent ELI's have better response rates.

To publish product indexes it is necessary to have several ELI's sampled in each stratum. Those strata in which only a few companies trade are filled up to specified levels with certainty selections from the "important" companies in these strata. The company burden is decreased by one whenever a certainty pick is made to insure that the total number of prices requested from a company is not too large.

At this point a single random start systematic PPS sample of ELI's is chosen for each company. The "relative-prob" (company-stratum prob divided by the max-prob) of each company-stratum is distributed among the ELI's in proportion to their dollar value. This "ELI-prob" is the measure of size in the second stage. The company burden, reduced by the number of certainty picks, is the number selected. This sample design has a fixed sample size, thus allowing control on the number of companies to be visited. The expected number of ELI's per stratum is calculated using

$$E(n_h) = \sum_{j=1}^{N} \sum_{i \in h} P_{1j} * P_{2hij} \qquad (1)$$

where P_{1j} = the first-stage probability of selection of company j

$$= \begin{cases} 1 & \text{for certainty companies} \\ \dfrac{n^* * \max p_j}{\sum_{k \in S_p} \max p_k} & \text{for probability companies} \end{cases}$$

and P_{2hij} = the second-stage probability of selection of company j - ELI i in stratum h

$$= \begin{cases} 1 & \text{for certainty selections} \\ \dfrac{p_{hij} / \max p_j}{I_{j2}} & \text{for probability selections} \end{cases}$$

where N = number of companies in the sampling frame,

n^* = sample size decreased by number of companies selected with certainty

$\max p_k$ = max-prob for company k,

S_p = universe of non-certainty companies,

p_{hij} = proportion of stratum h's dollar value in company j - eli i,

and $$I_{j2} = \sum_{i \in S_{jp}} \frac{p_{hij}}{\max p_j * B_j}$$

where S_{jp} = set of non-certainty ELI's in company j,

and B_j = number of probability selections in company j (burden reduced by the number of certainty picks)

Formula (1) with an additional subscript for consistency rank (p_{hijCR}) is used to calculate the expected number of ELI's in each stratum by consistency rank. The expected number of sample selections per company is the company burden calculated for second stage selection.

The results obtained using the two sets of test data were as expected. More than 60% of the expected ELI's were consistent with an average number of ELI's per company between four and five.

The final process in the current sampling design is the process of selecting an actual product to be priced. This stage was not used in comparing the alternative designs because it would remain the same for each design. The process of subsampling within an ELI consists of successive stages of disaggregation into subclasses, assigning a measure of size to each subclass, and making a selection of a subclass with probability proportionate to the assigned measure of size. The process ends when either a priceable item is reached or further disaggregation appears impossible.[6] In the latter case, a priceable item within the last subgroup is selected judgmentally by asking the respondent for a typical item. For each stage disaggregation is done according to any classification natural to the respondent for the product category, not constrained by any official classification system. The measures of size used are in terms of the percentage of value within the class being disaggregated.

LAHIRI - DESIGN 2

The first alternative design considered is a theoretical approach first presented by Lahiri upon which our present design is based. This design is also a two stage sampling procedure with control over the number of companies sampled. The first stage of selection is identical to the first stage of selection in the present design described above.

The second stage is a selection of ELI's within each company. The selection is performed using a single start systematic PPS selection. However it differs from the present design in that the interval is fixed at one, the measure of size of each ELI is $p_{hij} / \max p_j$, and the number of hits per company is a variable. Every establishment will have exactly one ELI selected in its max-prob stratum and at most one ELI selected in each additional stratum in which it deals.

One of the features about the Lahiri design is the correlation of the expected number of hits in a stratum to the sample size. Using the following formula, it can be shown that once the sample size is chosen, the expected number of hits in each stratum is a constant dependent only on n and the sum of the "max-probs".

$$E(n_h) = \sum_{j=1}^{N_h} \sum_{i \in h} P_{1j} * P_{2hij}$$

$$= \sum_{j=1}^{N_h} \sum_{i \in h} \frac{n * \max p_j}{\sum_{k=1}^{N} \max p_k} * \frac{p_{hij}}{\max p_j}$$

$$= \sum_{j=1}^{N_h} \sum_{i \in h} \frac{n * p_{hij}}{\sum_{k=1}^{N} \max p_k}$$

$$= \frac{n}{\sum_{k=1}^{N} \max p_k} \sum_{j=1}^{N_h} \sum_{i \in h} p_{hij} = \frac{n}{\sum_{k=1}^{N} \max p_k}$$

because $\sum_{j=1}^{N_h} \sum_{i \in h} p_{hij} = 1$.

where N_h = number of companies in stratum h,
n = sample size,
and all other terms as defined before.

In our analysis we made n dependent on $E(n_h)$.

$$n = E(n_h) * \left(\sum_{k=1}^{N} \max p_k \right)$$

We chose $E(n_h)$ based on our response rates, and used the above equation to determine the appropriate sample size.

This value n assumes that there are no certainties or that we allowed multiple hits. Neither was the case. However, we decided to use this number since 1) handling certainties would have to be done on a stratum by stratum basis and would result in a different n_h for each stratum, and 2) program considerations make multiple hits impractical.

Using this n as the sample size, we selected the first stage certainty units from each set of data. Then the following formula could be used to determine the expected number of company-ELI's in each stratum.

$$E(n_h) = \sum_{j\in S_{ph}} \sum_{i\in h} \frac{n^* * \max p_j}{\sum_{k\in S_p} \max p_k} * \frac{p_{hij}}{\max p_j} + \sum_{j\in S_{ch}} \sum_{i\in h} 1 * \frac{p_{hij}}{\max p_j}$$

$$= \frac{n^*}{\sum_{k\in S_p} \max p_k} \sum_{j\in S_{ph}} \sum_{i\in h} p_{hij} + \sum_{j\in S_{ch}} \sum_{i\in h} \frac{p_{hij}}{\max p_j} \qquad (2)$$

where S_{ph} = set of non-certainty companies in stratum h,

S_{ch} = set of certainty companies in stratum h,

and all other terms as defined before.

Formula (2) with an additional subscript for consistency rank is used to determine the expected number in each stratum by consistency rank. The distribution of hits per company is computed using only the second stage of selection (i.e. assuming the company is selected). The formula is

$$E(m_j) = \sum_{i\in j} \frac{p_{hij}}{\max p_j}$$

where m_j = number of hits for company j.

The results of our test showed that more than 70% of the expected ELI's were consistent. The Lahiri methodology had the smallest number of expected hits per company but had several companies with eight or more expected hits, causing a possible burden problem.

INDEPENDENT SELECTION - DESIGN 3

Independent selection of ELI's within strata is very different from the two preceding designs. In this case, control over the number of companies is sacrificed in order to guarantee sufficient coverage in each category to be published. The number of hits in each stratum can be specified and the method of selection can be varied. The type of selection in each stratum was systematic PPS with the measure of size being the company-ELI's dollar value. The expected number of hits in a company was calculated using $\sum_h \sum_i (p_{hij} * n_h)$ where p_{hij} is the proportion of stratum h's dollar value in company j - ELI i. The expected number in each consistency rank is

$$E(n_{CR}) = \sum_h \sum_i \sum_j p_{hijCR} * n_h.$$

Finding the expected number of companies in a sample is more complicated because 1) there can exist more than one ELI in a company in a particular stratum, and 2) we use a systematic PPS without replacement sampling algorithm. Therefore, the probability of selecting a company in a stratum is dependent upon the sort of its ELI's. A company would have its maximum chance of selection if all its ELI's were sorted together, and it would have its minimum chance if all its ELI's were located in the same part of the sampling interval. In order to find a company's chance of selection in a stratum (P_{jh}) we overlaid one interval on top of another and calculated the proportion of this one interval which was covered by any ELI in the company. We then determined the company's chance of not being selected in the stratum using $Q_{jh} = 1 - P_{jh}$. Since all the strata are independent, the probability of not selecting a company is

$$Q_j = \prod_{h=1}^{L} Q_{jh} \quad \text{(where L = number of strata) and}$$

the probability of selecting a company is

$$P_j = 1 - Q_j.$$

Simply summing the P_j over all companies gives you the expected number of companies in the sample.

As mentioned before, one advantage of independent selection which has not been fully explored is the ability to allocate the number of hits to each stratum. However, this flexibility does exist and is of major significance when considering independent selection.

The results from our test of independent selection were very good. The expected number of companies in the sample was similar to the preceeding two methods. This was not intuitively expected. In addition the expected number of consistent ELI's made up more than 76% of the sample. This was higher than any of the other methods.

One disadvantage was the substantial number of companies which had an expected number of selected ELI's less than two. This is not very cost effective since a large portion of the program's funds goes towards each company visit. Another disadvantage is that some companies had too many ELI's. If selecting within a stratum solved the problem, then the only result would be the loss of a few quotes. However, if you had to select across strata, then the independence of the design would be lost.

COMPANY SELECTION BY PRODUCT - DESIGN 4

This design is a two stage procedure in which the first step is the selection of companies and the second step is the selection of products within the companies. The first stage is identical to independent selection except that when one ELI in a company is selected, the company is selected for every ELI in which it traded during the reference period. Therefore the probability of selecting a company is

$$P_j = 1 - \prod_{h=1}^{L} (1 - P_{jh})$$

which is the same as in independent selection. The expected number of hits in each stratum after the first stage is $E(n_h) = \sum_j \sum_i p_{hij}$

where $p_{hij} = P_j$. The expected number in each consistency class is $E(n_{CR}) = \sum_h \sum_i \sum_j p_{hijCR}$

where $p_{hijCR} = P_j$.

The selection procedure for products within companies was the same one currently used in our present design. A burden was computed for each company using the algorithm in method 1. The chance of selection of each ELI in a company (also defined in method 1) was P_{2hij}. The final value for $E(n_h)$ is calculated as follows:

$$E(n_h) = \sum_j \sum_i P_j * P_{2hij}.$$

and $E(n_{CR})$ is calculated as follows:

$$E(n_{CR}) = \sum_h \sum_j \sum_i P_j * P_{2hijCR}.$$

This design has the advantage of independent selection in that you can allocate to and thereby guarantee sufficient coverage in each stratum. Also you can vary the method of selection if the situation warrants. It also has the advantage of our present design where you pick up everything in a company for a minimal cost. Its disadvantages include losing control over the number of companies in the sample and having to subsample certain companies in order to reduce their burden. In addition, some of the estimation and variance calculation properties of independent selection are lost.

ANALYSIS

In our analysis we assumed that we wanted an average of 40 company-ELI's per stratum. Using the Lahiri formula we calculated sample sizes of 1,621 companies for the Fifth Importer data and 1,289 companies for the Sixth Importer data. These sample sizes were used for our present design as well as for the Lahiri design. We determined the expected sample size for Designs 3 and 4 by setting the number of expected quotes per stratum equal to 40 so as to make comparison with Lahiri more meaningful. For the Fifth Importer data, this led to an expected sample size of 1,570 companies. The expected number of companies for the Sixth Importer data was 1,528. As mentioned before, the expected sample size is based on the method by which the ELI's are sorted. These sample sizes are based on a sort by consistency rank, dollar value, and A-number within each stratum. We calculated the expected number of companies based on a sort of the ELI's by company within strata to be 1,596 and 1,555 for the Fifth and Sixth Importer data, respectively.

The second factor in our analysis is the expected number of company-ELI's. Table 1 shows the expected number of ELI's for each design. We see that Designs 1 and 4 provide about three times as many company-ELI's in the sample as the other methods.

Table 2 shows the expected number of ELI's by consistency rank. Table 3 shows the expected percentage of ELI's by consistency rank for each design. Of the four designs, the third had the highest percentage of consistent ELI's while Design 4 had the lowest. However Design 4 had the largest number of consistent ELI's for the Sixth Importer data and Design 1 had more consistent ELI's for the Fifth Importer data.

Another factor to consider when comparing these designs is the expected number of ELI's per company. This is calculated as the ratio of the number of ELI's to the number of companies in the sample. As shown in Table 1 the first and fourth designs had approximately the same ratio of ELI's per company and a much larger ratio than the two other designs.

Table 4 shows the distribution of the expected number of ELI's per company for the two data sets. In the independent selection model (Design 3) there is only one stage of selection so the expected number of ELI's for each company is the sum of the probabilities of selection of the company's ELI's. Thus there can be companies with less than one expected ELI. For the other three designs, there are two stages of selection, company selection and selection of ELI's within each company selected in the first stage. The expected number of ELI's per company for these methods is calculated based on the assumption that the company is selected during the first stage and does not include the first stage probability of selection. Presently we allow each company to be selected for at most seven ELI's in a sample, so any company with more than seven ELI's would have to be subsampled to reduce its burden to seven or less. As seen in Table 4, there were several companies with more than seven ELI's for both the second and third designs. The first and fourth methods are identical in the second stage of selection so their distributions of expected ELI's per company are identical.

CONCLUSIONS

Methods 1 and 4 provide samples with about three times the expected number of ELI's as Designs 2 and 3 while retaining the same number of companies to be visited. The second and third designs also contain companies which would require subsampling to reduce their burden. While the theoretical Lahiri (Design 2) and independent selection (Design 3) models have a greater percentage of consistent ELI's, we feel that their lower number of ELI's is not sufficient to meet the publishability requirements of the IPP.

The present technique (Design 1) and the company selection by product model (Design 4) provided very similar results for our test data. Since a major factor in our analysis is the number of companies in the sample, we were very interested in the expected number of companies in Design 4. For our test data, the sample sizes for both designs were very similar even though the sample size was not controlled for the fourth design and it was controlled for the first method. The ability to control the sample size while obtaining a large number of ELI's for the sample without a burden problem has lead us to conclude that our present design is the best fitted to the goals of the International Price Program.

REFERENCES

1/ BLS Handbook of Methods, United States Department of Labor, Bureau of Labor Statistics Bulletin 1910, 1976.

2/ Lahiri, D.B., A Method of Sample Selection

Providing Unbiased Ratio Estimates, Bulletin of the International Statistical Institute, Volume XXXIII Part II.

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

4/ Schedule A Commodity by Country, Bureau of the Census, FT135, October 1978.

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

6/ Report on Tasks III, IV, and V: Design Alternatives, Field Test Plan, and Materials, Westat, Inc., July 21, 1977.

TABLE 1

Comparison of Alternative Sample Designs

Section 7 - Fifth Importer Data

	Number of ELI's	Number of Companies	Average Company Burden
Design 1	7258	1621	4.477
Design 2	2267	1621	1.398
Design 3	2749	1570	1.751
Design 4	7113	1570	4.531
Universe	104090	29459	3.533

Section 8 - Sixth Importer Data

	Number of ELI's	Number of Companies	Average Company Burden
Design 1	6336	1289	4.916
Design 2	2016	1289	1.564
Design 3	2240	1528	1.466
Design 4	7736	1528	5.063
Universe	254859	63489	4.014

TABLE 2

Distribution of Expected Number of ELI's by Consistency Rank

Section 7 - Fifth Importer Data

	Consistency Rank			
# of ELI's	1	2	3	4
Design 1	1981	918	1483	2876
Design 2	395	206	412	1254
Design 3	429	207	442	1671
Design 4	2001	905	1451	2756
Universe	65877	14615	12869	10729

Section 8 - Sixth Importer Data

	Consistency Rank					
# of ELI's	1	2	3	4	5	6
Design 1	1644	584	525	460	832	2291
Design 2	256	118	136	130	274	1102
Design 3	278	117	136	127	278	1304
Design 4	2072	768	694	593	1056	2553
Universe	165020	33889	15026	12493	13619	14812

TABLE 3

Distribution of Expected Percentage of ELI's by Consistency Rank

Section 7 - Fifth Importer Data

	Consistency Rank			
% of ELI's	1	2	3	4
Design 1	27.3	12.7	20.4	39.6
Design 2	17.4	9.1	18.2	55.3
Design 3	15.6	7.5	16.1	60.8
Design 4	28.2	12.7	20.4	38.7
Universe	63.3	14.0	12.4	10.3

Section 8 - Sixth Importer Data

	Consistency Rank					
% of ELI's	1	2	3	4	5	6
Design 1	25.9	9.2	8.3	7.3	13.1	36.2
Design 2	12.7	5.8	6.7	6.5	13.6	54.7
Design 3	12.4	5.2	6.1	5.7	12.4	58.2
Design 4	26.8	9.9	9.0	7.7	13.6	33.0
Universe	64.8	13.3	5.9	4.9	5.3	5.8

TABLE 4

Distribution of Expected ELI's per Company

Section 7 - Fifth Importer Data

Expected	Design			
ELI's/Company	1	2	3	4 **
0 - 1 *	15,269	15,269	28,890	15,269
1 - 2	8,161	12,533	354	8,161
2 - 3	1,901	1,362	103	1,901
3 - 4	1,041	233	34	1,041
4 - 5	1,214	42	20	1,214
5 - 6	1,387	12	11	1,387
6 - 7	486	4	10	486
7 - 10		3	15	
10 - 15		0	13	
15 - 20		0	5	
over 20		1	4	

Section 8 - Sixth Importer Data

Expected	Design			
ELI's/Company	1	2	3	4 **
0 - 1 *	30,105	30,105	63,067	30,105
1 - 2	17,991	28,359	309	17,991
2 - 3	4,566	3,737	59	4,566
3 - 4	2,671	840	20	2,671
4 - 5	3,236	264	19	3,236
5 - 6	3,520	93	2	3,520
6 - 7	1,400	41	1	1,400
7 - 10		40	6	
10 - 15		9	3	
15 - 20		0	0	
over 20		1	3	

* For Design 3 there can be less than one ELI per company but for Designs 1, 2, and 4 there must be at least one ELI per company.

** The second stage of Designs 1 and 4 is identical so the expected number of ELI's is the same.