

SUPPORTING STATEMENT FOR REQUEST FOR EMERGENCY ACTION–PART B
UNITED STATES INTERNATIONAL TRADE COMMISSION QUESTIONNAIRE

China: Effects of Intellectual Property Infringement and
Indigenous Innovation Policies on the U.S. Economy

B. Collection of Information Employing Statistical Methods

1. Response universe and sampling strata

The potential respondent universe includes all companies that conduct business in the United States and have primary NAICS codes in particular areas that are considered by USITC to have a risk of loss due to IPR infringement practices by Chinese entities. These industries were identified by USITC staff.

A stratified sample based on a two-step stratification process is being implemented for this project. Firms are first sub-divided into the following categories as determined by USITC staff:

1. Industries with high potential for losses due to counterfeiting, other than 4-6 below
2. Industries with high potential for experiencing patent infringement , other than 4-6 below
3. Industries with high potential for experiencing copyright infringement, other than 4-6 below
4. Software publishing
5. Professional Services
6. Pharmaceutical & Other Chemicals
7. Firms in other industries with high potential for experiencing other types of IP infringement either because they have valuable trademarks or FDI in China

Then, within each primary category, firms are then further subdivided by NAICS code, which constitutes the level at which the sample is being selected. Within each stratum, firms with relatively high levels of FDI, revenue, or trademark value are sampled at 100%, while the remainders of firms are selected from randomly. The number of firms in each strata and the sample size is provided in Table 1 (below).

TABLE 1 Population and Sample Count in Considered Sectors

Type of Infringement	NAICS	Population Count	Sample Count	
Counterfeit	Breweries, Wineries, Distilleries	31212_4	4,054	102
	Tobacco Manufacturing	3122	502	100
	Apparel Manufacturing	315	21,839	117
	Footwear Manufacturing & Other Leather & Allied Product	3162&9	4,062	102
	Watch, Clock, and Part Manufacturing	334518	313	100
	Jewelry and Silverware Manufacturing	33991	8,223	104
	Game, Toy, and Children's Vehicle Manufacturing	33993	5,006	102
Other Patent	Machinery Manufacturing	333	78,586	337
	Computer and Electronic Product Manufacturing ¹	334	41,627	363
	Semiconductor and Other Electronic Component Manufacturing	3344	8,326	168
	Electrical Equipment, Appliance, and Component Manufacturing	335	16,804	171
	Motor Vehicle Equipment Manufacturing	3361_3	12,848	182
	Aerospace Product and Parts Manufacturing	3364	3,919	123
	Medical Equipment and Supplies Manufacturing	3391	21,913	158
Other Copyright	Newspaper, Periodical, Book, and Directory Publishers	5111	58,303	480
	Motion Picture and Video Industries	5121	54,810	292
	Sound Recording Industries	5122	995	100
	Television Broadcasting	51512	2,208	228
	Internet Publishing and Broadcasting and Web Search Portals	51913	56	56
Software Publishers	Software Publishers	5112	538	226
Professional Svcs	Computer Systems Design and Related Services	5415	129,995	677
	Research and Development	5417	53,091	516
Pharm & Other Chems	Chemical Manufacturing ¹	325	17,797	260
	Pesticide, Fertilizer, and Other Agricultural Chemical Manufacturing	3253	1,228	105
	Pharmaceutical and Medicine Manufacturing	3254	5,808	124
Other	Other		382	382
TOTAL			441,839	5,675
¹ Firms in these categories have FDI in China for which only 3-digit NAICS information was available at the time of writing.				

The sample size of 5,675 is the number of surveys that will be sent out and is based on what is needed for a statistically significant response, given historical response rates. Results of similar past surveys, we expect the response rate to range from 25–45 percent. Given a response rate in this range, the number of surveys received from the sampled companies ranges from 1,250 to 2,250.

2. Collection of information

a. Statistical methodology for stratification and sample selection

The goal of the stratification scheme is to develop a set of strata that minimize variance within strata in terms of the amount and type of losses experienced due to IPR infringement. Because no pro-forma reliable data exist on the expected total amount of IPR infringement experienced by different industries, the stratification scheme was based on the best judgment of industry and ITC experts regarding which types of industries would be more likely to experience similar amounts and types of losses from IP infringement. Therefore, firms were stratified according to the type of IP infringement likely to be experienced and by industry type.

The sampling frame for this project was created from the ORBIS database, a proprietary database of information on all known firms. Firms were categorized by their primary NAICS code, which reflected the industry sector in which the largest portion of the firms' revenue was derived.

When sampling within strata, the ITC wants to ensure that particular types of firms are sampled at 100%, as these firms are extremely likely to have levels of losses from IP infringement that are highly divergent from other firms.

1. Firms with “Top 100” trademarks as indicated by Interbrand and supplemented by ITC staff.
2. Firms with FDI in China as indicated in the Orbis database.
3. From the remaining list of firms (after excluding firms that satisfy the first two criteria above), does the company rank in the top 100 in terms of revenue within each of the categories (1)-(6) identified in Question 1?

In terms of sample identification, the Strata Top 100, FDI, & Top 100 Trademark companies are sampled with certainty. In other words, if a company has any of these three distinctions it is guaranteed to be part of the final sample. Firms that do not have one of these three “distinctions” are sampled at random.

The maximum sample of 5,675 is based on a response needed to differentiate answers from sectors with statistical significance. After ensuring that a minimum number of firms are sampled in each of the IP sensitive industries, the methodology for allocating the remaining firms is based on the Neyman method, which is designed to provide the optimal mix for the sample allocation when using a stratified random approach. The driver of allocation using this scheme is the product of each stratum’s remaining population and standard deviation of its revenue. By doing so, it forces more sample to stratum that have a higher product with the end goal of lowering the within stratum subsequently improving the overall precision of the analysis. In the case of the sound recording industry, firms will be chosen from the membership list of the Recording Industry Association of America (RIAA), which has superior coverage to Orbis.

b. Estimation Procedure

In order to produce population estimates and precision statistics about the estimate, the following equations will be used:

The formula used to estimate the population attribute of interest is found in Equation 1. The precision statistics about the estimate are found in Equations 2-4. Per standard notation, the total estimate from a stratified random sample τ_{st} is shown in Equation 1:

$$\text{Equation 1}$$

$$\tau_{st} = \sum_{h=1}^L N_h \bar{y}_h$$

where h equals an individual stratum, N_h equals the population of stratum h , \bar{y}_h equals the average of the attribute of interest of the sampled items in stratum h . For example, \bar{y}_h would represent the average amount of revenue loss within each NAICs code.

Its variance estimate is per Equation 2:

Equation 2

$$\text{var}(\tau_{st}) = \sum_{h=1}^L N_h(N_h - n_h) \frac{s^2}{n_h}$$

where s^2 equals the standard deviation of the attribute of interest within stratum h . Its standard error is per Equation 3:

Equation 3

$$\text{Standard error} = \sqrt{\text{var}(\tau_{st})}$$

c. Degree of accuracy needed for the purpose described in the justification

The sample size for this survey is constrained by USITC resources for processing the completed questionnaire. Given that constraint, it is expected that it will be feasible in most cases to distinguish the responses in categories (1)-(7) within a 95 percent confidence, both for the binary questions and for questions requiring responses in U.S. dollars. We also expect to be able to report dollar value totals with 95 percent confidence to within a coefficient of 0.1 for the full sample and 0.5 for categories (1)-(7) and in some cases for individual sectors. This degree of confidence is sufficient for the purposes described in the justification.

d. Unusual problems requiring specialized sampling procedures

No unusual problems were encountered.

e. Any use of periodic (less frequent than annual) data collection cycles to reduce burden.

This data collection is currently only intended to occur once, therefore will not be repeated on any periodic basis.

3. Response rates

Commission staff will employ several techniques to increase the response rates of questionnaire recipient firms. Selected recipients will receive up to 4 notices. Firms will first receive a letter notifying them that their firm was selected to fill out the Commission survey and that it will arrive shortly. Firms will then receive a letter directing them to complete the Commission survey. Prior to the response deadline, firms that have not responded will receive a reminder post card of the impending submission deadline. Once the submission deadline has passed, firms that still have not responded will receive an additional letter directing them to fill out the questionnaire and submit it. Each of these communications will include a phone number and email address of a person that can help firms with filling out the questionnaire or answer their questions regarding the survey and/or study. Commission staff may also contact firms directly, via phone or email, to urge them to complete the survey and to answer any questions they may have regarding this information collection or study in general. Commission staff may also contact firms, via phone or email, to correct information or fill in incomplete responses, or solicit additional information about their response. The burden associated with follow up calls or emails is included in the total response burden amount.

Irrespective of the response rate of the survey, once the survey response period has been closed, the ITC will conduct an analysis of survey responses in order to detect and correct for any non-response bias contained in the survey results. Analysis for non-response bias seeks to answer two questions: Do the characteristics of those who answer the survey differ from those that do not answer the survey? And, are those characteristics associated with the survey's outcomes of interest?

In order to examine the impact on non-response bias, the team will first examine conditional response rates for groups of firms based on characteristics available in the data frame that are hypothesized to

impact outcomes of interest. These may include variables such as firm size, size of interests in China, industry, primary product or service, NAICS or SIC code, or other variables. Any differences in response rates (whether statistically significant or nominal differences of a large magnitude) can be further investigated through logistic regression analysis, using firm characteristics as predictors, and whether or not a recipient responded to the survey as a binary outcome. If the results of the regression show that none of the characteristics influences the propensity of a survey recipient to respond to the survey in a statically significant way, then the team will make the assumption that non-responses are randomly distribute in the population, and that no correction for non-response bias is required.

However, if the results of the logistic regression indicate that one or more of the characteristics investigated above affects the propensity of a survey recipient to respond to the survey, then those characteristics should be examined to determine whether they are associated with differences in the outcome variables under study across the dataset of survey responses collected. This could be investigated using various multivariate methods, using firm characteristics as predictors, and the survey variable of interest as the outcome. If any sources of non-response bias are found, they can be controlled for by the development of weights, which can then be used in concert with weighting based on population stratification, in the extrapolation of results to the entire population.

4. Test for procedures or methods

Once the samples within each of the strata described above have been completed, the Commission will conduct several diagnostic tests of the sample including comparing means and variances between the sample and the population for each of the strata in order to identify any potential biases or problems with the sampling technique. These diagnostic tests will not require any communication with potential respondents.

Besides the previously described pretesting/field testing with firms, the questionnaire has been extensively reviewed within the Commission. Industry analysts and economists who will prepare the report have reviewed the document and made recommendations to improve the content of the questionnaire so that it will include the information needed to adequately answer questions addressed in this study while imposing a minimum burden on the responding businesses. A survey consultant reviewed the questionnaire to assure that it is clear and in a form that is likely to elicit the desired responses.

5. Contact information

Collection and analysis of the data will be the responsibility of the Office of Economics and the Office of Industries within the Commission. Project leader, Alexander Hammer, of the Office of Economics, can be contacted at 202-205-3271, deputy project leader, Jeremy Wise, of the Office of Industries, can be contacted at 202-205-3190, and lead economist for this study, Michael Ferrantino, can be contacted at 202-205-3241. In addition to numerous Commission staff consulted on the statistical aspects of the questionnaire, Commission staff also worked with Albert Lee and his staff at Summit Consulting, a survey design and data analysis consulting firm. Mr. Lee may be contacted at (202) 407-8300.