SUPPORTING STATEMENT
UNITED STATES INTERNATIONAL TRADE COMMISSION QUESTIONNAIRE

Investigation No. 332-586

*Foreign Censorship Part 2: Trade and Economic Effects on U.S. Businesses*

Part B-Collection of Information Employing Statistical Methods

## 1. Response universe, sample sources, and sampling strata

### Survey objectives

In a letter dated April 7, 2021, the Committee on Finance of the U.S. Senate (Committee) directed the Commission to produce two reports that identify various censorship practices and analyze the trade and economic effects of such practices in key foreign markets. The Committee specified that the Commission's report be based on a review of available data, including a survey of U.S. businesses.

### Respondent universe

The respondent universe includes U.S. businesses that have operations in China or have filed a patent in China. The USITC has identified 5,573 such businesses for its sampling frame, of which 3,790 will be sampled. The sampling unit is the business rather than the establishment.

The potential respondent universe represents the sum of businesses, net of duplicative records, identified in the Bureau van Dijk's Orbis database as having operations in China in the previous 30 months (as of March 2021) and spending activity in the previous 12 months. Businesses that operate in China are those in the database that meet at least one of the following criteria: 1) foreign shareholders located in China (at least 10%); 2) subsidiaries located in China (at least 10% ownership); 3) involved in FDI projects in China; 4) involved in mergers and acquisitions in China; or 5) filed a patent in China.

### Sample design

Survey respondents will be selected through a stratified random sampling methodology that stratifies businesses through a combination of industry, size, and whether the business has operations in China or has only filed a patent in China. Stratification groups businesses that are likely to respond similarly into stratum and the random sample will help reduce bias in the survey results. The stratification plan is based on three sets of expectations. 1) Censorship is likely to impact businesses in service industries more than manufacturing, agriculture, or mining, so staff will stratify the population by industry group (Services only, Manufacturing only, or Services and Manufacturing) using primary and secondary North American Industry Classification System (NAICS) codes. 2) Businesses with operations in China (as defined in respondent universe, items 1 through 4) are expected to respond differently than businesses that have only filed patents in China (respondent universe, item 5), so staff will separate these companies. 3) Small and medium enterprises (SMEs) are expected to experience and respond to censorship acts, policies, and practices differently than large businesses with more resources to comply with such policies, so staff will stratify by size of business.

Table 1 presents the size cut-off criteria used to stratify into SMEs and large businesses by industry group. Table 2 presents the sampling frame—the population of businesses in each stratum. Table 3 presents the sample size for each stratum.

Based on results of past surveys conducted by the Commission for other investigations, the average response rate for similar surveys has been approximately 45 percent. Staff believe the response rate for this survey collection may be lower than average because sampled businesses may be less inclined to respond due to the sensitivity of foreign censorship. Staff will pull a sample of 3,790 businesses based on our minimum sample size calculation per stratum (Table 3). Responses in previous and ongoing Commission surveys have not differed significantly by business size or across industries. Thus, a uniform response rate has been assumed for all strata.

**TABLE 1** Size cutoff criteria for size-based stratum by industry group

| Industry group | SME | Large |
|---|---|---|
| Services | <= $1 billion in revenue | > $1 billion in revenue |
| Manufacturing | <= 500 employees | > 500 employees |
| Both | <= $1 billion in revenue | > $1 billion in revenue |

**TABLE 2** Sampling frame—population per stratum

| Industry group | Operations | | Filed Patents Only | | Total |
|---|---|---|---|---|---|
| | SME | Large | SME | Large | |
| Services | 412 | 230 | 1,725 | 137 | 2,504 |
| Manufacturing | 148 | 240 | 655 | 261 | 1,304 |
| Both | 266 | 211 | 1,195 | 93 | 1,765 |
| Total | 826 | 681 | 3,575 | 491 | 5,573 |

Note: The manufacturing industry group includes agriculture and mining NAICS codes. “Both” includes businesses associated with both manufacturing and services NAICS codes.

A power analysis was conducted to determine the minimum sample size needed per stratum in order to produce statistically valid results, with a 90% two-sided confidence interval, based on the size cutoff criteria listed in table 1. The results of the power analysis, sample size per stratum, are shown in table 3.

**TABLE 3** Sample size per stratum

| Industry group | Operations | | Filed Patents Only | | Total |
|---|---|---|---|---|---|
| | SME | Large | SME | Large | |
| Services | 389 | 230 | 942 | 137 | 1,698 |
| Manufacturing | 102 | 240 | 378 | 261 | 981 |
| Both | 227 | 211 | 580 | 93 | 1,111 |
| Total | 718 | 681 | 1,900 | 491 | 3,790 |

**2. Collection of information employing statistical methods**

**a. Statistical methodology for stratification and sample selection**

A stratified sample is being implemented for this collection. The goal of the stratification scheme is to develop a set of strata that minimizes the variance of responses (such as level of employment and industry group) within each stratum. Because no pro-forma reliable data exist on the size and scope of the number of businesses that are impacted by foreign censorship in China, the classification of size within industry groups for the stratification scheme were estimated based on previous Commission surveys.

Sampled businesses will be randomly selected by stratum; each business will have the same probability of being selected as others in its stratum, based on the size of the stratum it is assigned to.

**b. Estimation Procedure**

Survey estimates will be based on weighted data. The weighting procedure will incorporate a sample selection weight, a nonresponse adjustment factor, and if necessary, a poststratification weighting factor. There is an equal probability of selection within each stratum.

- *Sample selection weighting*: Because the sampling rates are based on three criteria, as discussed above, the selection weight factor will account the probability of selection within a particular industry, size, whether they have operations or only filed a patent, and any oversampling of businesses.
- *Nonresponse adjustment*: The nonresponse adjustment factor is designed to attenuate bias due to differential response rates. This adjustment will be calculated using business characteristics, if warranted. See the section below on accuracy and reliability of information collected for further discussion.
- *Poststratification weighting*: If necessary, a poststratification weighting factor will be used to attenuate bias due to sample frame noncoverage, overcoverage, or omissions. However, there are no other data available that limits companies by their activities in other countries; therefore, the Commission is not likely to conduct post-stratification weighting.

The general weighting formula can be represented as

$$W_h = S_h \times NR_h \times PS_h , \quad (1)$$

where $S_h$ is the sample selection weight for stratum h, $NR_h$ is the nonresponse adjustment factor for stratum h, and $PS_h$ is the poststratification weight of stratum h. $W_h$is the weight applied to all observations in stratum h. This formula may be adjusted to include a business-specific weighting component if non-response is determined to be related to factors aside from the factors used to design the strata.

Standard estimation procedures will be used as in Heeringa et al (2010).[1] For example, the formula used to estimate the population attribute of interest is found in equation 2. Per standard notation, the total estimate for industry $k$, $\tau_k$, from a stratified random sample, is given by

[1] S. Heeringa, B. West, and P. Berglund, *Applied Survey Data Analysis*, CRC Press, 2010.

$$\tau_k = \sum_{h=1}^{L} N_h \bar{y}_h, \quad (2)$$

where *h* denotes an individual stratum, $N_h$ equals the population of stratum *h*, and $\bar{y}_h$ equals the average of the attribute of interest of the sampled items in stratum *h*. For example, $\bar{y}_h$ could represent the average amount of revenue within each stratum.

The variance estimate for sampling without replacement is given by

$$var(\tau_k) = \sum_{h=1}^{L} N_h (N_h - n_h ¿) \frac{s^2}{n_h} ¿ \quad (3)$$

where $s^2$ equals the standard deviation of the attribute of interest within stratum *h*, and $n_h$ is the sample size for stratum *h*.

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

A sample size of 3,790 is needed to achieve estimates of +/- 5 percent at 90 percent confidence. It is expected that it will be feasible to produce statistically significant results for most survey items at the aggregate level at a 90 percent confidence level, both for the continuous and categorical variables. For example, table 4 provides the maximum margin of error for a binary question, given alternative response rates. Non-binary questions will have larger margins of error. However, based on other surveys with similar designs and weight adjustments, relatively small design effects (DEFFs), in the 1.4 to 1.8 range may be expected for most estimates. The large sampling rates in some strata will also reduce variance due to the finite population correction.

**TABLE 4** Margin of error for 90 percent confidence interval

| Measure | 20% response rate | 30% response rate | 40% response rate |
|---|---|---|---|
| Number of respondents | 750 | 1,125 | 1,500 |
| Standard error, percent | 1.83 | 1.49 | 1.29 |
| Margin of error, percent | 3.00 | 2.45 | 2.12 |

Note: This assumes a maximum margin of error of 50 percent for a binary question.

**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, and therefore will not be repeated on a periodic basis. As such, the total recurring annual cost burden is zero.

**3. Methods to maximize response rates and deal with non-response**

**a. Maximizing response rates**

Commission staff will employ several techniques to increase the response rates of questionnaire recipient businesses. Recipients will receive separate notices that (1) notify them that their business was selected for the survey, (2) direct them to complete the survey, and (3) remind them, if necessary, to complete the survey before the deadline. Once the submission deadline has passed, businesses that still have not responded will receive an additional reminder. Each of these communications will include a phone number and email address for the investigative team where businesses can receive help with filling out the questionnaire or have their questions regarding the survey and/or study answered. Commission staff may also contact businesses 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 businesses, via phone or email, to correct information or fill in incomplete responses, or solicit additional information about a response. The burden associated with follow up calls or emails is included in the total response burden amount.

In addition to pre-contact and follow-up, the questionnaire itself has been designed to be as clear and succinct as possible to gather the specific material requested by the Committee. (See discussion of testing below.) This clarity and brevity should reduce burden and improve response rates. The questionnaire will clearly point out that businesses are obligated by law to respond. Finally, the ability to access, fill out, and submit the survey electronically may also increase response rates.

**b. Accuracy and reliability of information collected**

The sample methodology has been designed to be as accurate and reliable as possible, based on Commission experience in past surveys. The sampling frame has been chosen to include businesses that have operations in China or have filed a patent in China.

Response rates in similarly scoped Commission surveys have recently been approximately 45 percent. The Commission will examine survey responses to detect and correct for any non-response bias. The team will first examine conditional response rates for groups of businesses based on characteristics available in the data frame that are hypothesized to impact outcomes of interest. These may include variables such as business size or industry. Any differences in response rates can be further investigated through logistic regression analysis, using business characteristics as predictors, and whether or not a recipient responded to the survey as a binary outcome. 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 will be examined to determine whether they are associated with differences in the outcome variables under study, across the dataset of survey responses collected. 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.

The Commission expects that all sampled information will yield reliable data that can be generalized to the universe studied.

**4. Tests of procedures or methods to minimize burden or improve utility**

The Commission sought comments on the questionnaire with industry representatives of several relevant industries through testing. These representatives provided feedback in areas such as availability

of data, reporting burden, product coverage and definitions, clarity of instructions, disclosure, and reporting format. See part A for the comments testers made, and the subsequent changes made to the questionnaire.

In addition to testing, the questionnaire has been made available for public comment. Notice of the draft questionnaire was published in the *Federal Register*. It has also been extensively reviewed within the Commission. Industry analysts and economists have reviewed the questionnaire to ensure it requests information needed to adequately answer questions posed in the study while imposing a minimum burden on the responding businesses.

The sampling methodology and procedures in this survey are similar to those in prior USITC survey work. Prior studies, for example, also have had populations drawn from Orbis; have also stratified by industry and size; and have used similar methods of survey distribution and data collection. Although the USITC has not specifically tested the methodology and procedures of the current Foreign Censorship survey, prior surveys have provided implicit tests of its practicability and utility.

**5. Contact information**

Collection and analysis of the data will be the responsibility of the Office of Analysis and Research Services, the Office of Economics, and the Office of Industries within the Commission. The project leaders for this investigation are Ricky Ubee, Shova KC, and George Serletis. The survey team can be reached by email at foreign.censorship@usitc.gov. If you prefer to contact them by phone, please call 202-780-1638.