



## Revision in the Sample Design of TANKAN using the Economic Census of Japan

Kyosuke Shiotani \*

Bank of Japan, Tokyo, Japan – [kyosuke.shiotani@boj.or.jp](mailto:kyosuke.shiotani@boj.or.jp)

### Abstract

The Bank uses a sample survey framework for the Tankan (Short-term Economic Survey of Enterprises in Japan). It randomly extracts samples from a population based on the "*Economic Census for activity*" conducted by the Ministry of Internal Affairs and Communications. "*Economic Census*" has been a series of newly created survey to reveal the realities of economic activity for all establishments and enterprises in Japan. Census surveys basic information for enterprises: name, location and so on, as well as accounting information: sales, expense and so on.

In the next revisions of sample enterprises, the Bank plans to improve the sampling methodology of Tankan by using this accounting information on "*Economic Census*," especially sales for each enterprise. Dividing the population into strata by the sales in addition to industry, the Bank can improve the effectiveness of stratified sampling and can achieve the statistical accuracy target with fewer samples.

**Keywords:** business survey; stratified sampling; economic census.

### 1. Introduction

The Tankan is a statistical survey conducted by the Bank of Japan to provide an accurate picture of business trends of enterprises in Japan, thereby contributing to the appropriate implementation of monetary policy. The Bank uses a sample survey framework for the Tankan, extracting sample enterprises from the population (excluding financial institutions) with capital of at least 20 million yen. To grasp actual economic conditions accurately, the Bank revises its population regularly in line with updates of the population base.

In March 2015 survey, the Bank made the first revision using the "*Economic Census*" as the population base for the Tankan survey. The "*Economic Census*" is jointly conducted by the Ministry of Internal Affairs and Communications and by the Ministry of Economy, Trade and Industry, and aims to provide a comprehensive overview of the actual business activities of establishments and enterprises in Japan and generate information on the population for various statistical surveys of establishments and enterprises. The "*Establishment and Enterprise Census of Japan*," a previous population base for the Tankan, has been integrated into the "*Economic Census*" from 2009.

The "Economic Census" consists of two surveys: the "Economic Census for Business Frame" and the "Economic Census for Business Activity." The "Economic Census for Business Activity" surveys the economic activities of establishments and enterprises, with its survey items including accounting information such as sales and capital investment amounts in addition to basic information such as business name, location, and number of employees. Since some Tankan survey items are also included in the accounting information collected by the "Economic Census", we can access detailed information for the target population. Insufficient time to prepare means that the regular revision conducted in March 2015 has not allowed for full utilization of information from the "Economic Census" in the Tankan. However, with a view to improving the accuracy and efficiency of the Tankan, it is very important to compare the Tankan results with the real values for the target population and discuss effective sampling methods for utilizing the information from the "Economic Census."

## 2. Evaluation of the Tankan results

We begin by checking the statistical accuracy of the Tankan. The Tankan has set accuracy targets for the ratio of the standard error to the level of sales (the error ratio) for actual and projected amounts of sales. Sales were chosen because they are an important survey item within the Tankan and are also useful as they correlate with the amount of business fixed investment to some extent.

Following the March 2010 revision, the information required for detailed calculation of the error ratio (the population mean and population variance of sales) could not be obtained from the *Establishment and Enterprise Census of Japan*, which does not provide sales data. Thus the Bank used the "Basic Survey of Japanese Business Structure and Activities" for the manufacturing sections and the "Census of Commerce" for the wholesale trade and retail trade sections, and estimated the population means and population variances based on the Tankan data for industries whose population information could not be obtained. However, the *Economic Census* enables us to obtain sales data for all enterprises in all sections, thereby facilitating calculation of the true value of the error ratio from the population mean and population variance.

Table 1: Error Ratio of the Sales by Industry and Scale

	Large		Medium-sized		Small		Sample size
	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.	
2010	1.5%	3.1%	2.6%	4.3%	2.3%	3.8%	11,684
2015	1.0%	3.0%	2.7%	4.6%	2.5%	3.9%	11,126

Notes: The Tankan categorized enterprises based on the size of capital: Large enterprises (capital with 1 billion yen and more), Medium-sized enterprises (capital with 100 million yen to less than 1 billion yen), and Small enterprises (capital with 20 million yen to less than 100 million yen). The Bank draws the sample enterprises to set the error ratio at less than 3 percent for the manufacturing sector and 5 percent for the nonmanufacturing sector for each size of enterprises.

Table 1 shows the error ratio for each of the six divisions by industry and scale at time of the March 2010 revision and the March 2015 revision. The error ratio for the manufacturing sector for Large enterprises decreases from 1.5 to 1.0. To achieve the accuracy target, the Bank divides strata into smaller segments by “industry,” by “capital” and by “the number of employees” (stratified sampling). Under the March 2015 revision, every stratum divided by the number of employees is set more flexibly compared to the previous revisions. This helps to improve statistical accuracy, particularly for the manufacturing sector for Large enterprises. For the same reason, the error ratio for the manufacturing sector for Large enterprises declines slightly from 3.1 to 3.0.

Error ratios increased for other industries and scales. This is because the population variances for sales based on the *Economic Census* data are higher than those used by the Bank in the previous revision. This outweighs the effect of setting each stratum flexibly. The *Economic Census* uses administrative record information (e.g. information on labor insurance, and information on commerce and corporate registration) to collect information on establishments and enterprises in addition to visual checks by enumerators, and has resulted in greater coverage of enterprises. Using the sales data from the *Economic Census*, the population variance increases particularly sharply for the real estate sector because there are more property management companies, which generate sales from real estate with relatively small numbers of employees, in the population. “The number of employees” that the Bank uses to divide enterprises into strata is an effective characteristic to decrease the error ratio of “sales” when it is highly correlated with sales. Dividing into groups by the number of employees, where this number is highly correlated with sales, we can make groups ordered by sales and thereby decrease the population variance of each stratum. This condition is satisfied for almost all sections. However, in the population based on the *Economic Census*, it is not satisfied for the real estate section and certain other sections with only weak correlation between sales and their number of employee.

Table 2: Sample Bias of the Tankan in Calendar Year 2011

	Large		Medium-sized		Small	
	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.
Sales	2.3%	11.3%	3.3%	19.8%	11.4%	26.4%
Fixed Investment	3.5%	9.8%	6.2%	20.2%	9.0%	13.7%

Next, using accounting information for all enterprises in the population, we check the sample bias of Tankan results compared with the true value for the population in the specific year that the *Economic Census* was conducted. Table 2 shows Tankan biases for sales and fixed investment excluding land purchasing expenses in calendar year 2011 with the March 2015 sample enterprises. Since the Tankan results are based on the Japanese fiscal year (from April to March), we use sales and fixed investment data for each enterprise from the *Economic Census* for comparison purposes instead of data from the Tankan survey. Thus the differences between the Tankan results and the population

values reflect sample bias only, and do not include any difference between Tankan responses and data reported in the *Economic Census* for the same company.

From Table 2, we can see that the Tankan estimates for the sum of sales and fixed investment were higher than the population values right across the board. Sample biases are found to be smaller for the manufacturing sector than for the nonmanufacturing sector, and smaller for Large enterprises than for Small enterprises. This is because the Tankan samples in each stratum are biased towards enterprises with large number of employees. The number of employees is highly correlated with sales and fixed investment on the whole, meaning that an upward bias in the number of employees causes upper biases for other items.

The main reason for the bias in the number of employees is perhaps a difference in the acceptance rate—that is, the ratio of the enterprises that agree to be in the Tankan sample to all enterprises that are asked to be in the sample—across enterprises with different numbers of employees. The population of the Tankan (enterprises with capital at least 20 million yen) includes enterprises with very few (less than ten) employees. The acceptance rate for such enterprises is lower than for enterprises with dozens of employees owing to their insufficient resources for accounting and responding the Tankan survey. Moreover, such enterprises are disproportionately represented in the nonmanufacturing sector for Small enterprises by comparison with the manufacturing sector for Large enterprises.

### 3. Improving the sampling methodology

How can these problems—higher error ratios and overestimation—be addressed? The key is to use sales data for all enterprises in the population (now available from the “*Economic Census*”) to divide the population into strata.

Table 3: Error Ratio based on the Strata Classified by Sales and Number of Employees

Items	Classified by	Large		Medium-sized		Small	
		Manuf.	Nonmanuf.	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.
Sales	Employees	1.0%	3.0%	2.7%	4.6%	2.5%	3.9%
	Sales	0.5%	0.8%	1.8%	3.1%	1.8%	1.6%
Fixed investment	Employees	3.0%	3.6%	6.5%	8.3%	10.1%	28.0%
	Sales	2.4%	3.1%	6.3%	8.2%	9.1%	11.8%

Table 3 compares error ratios based on strata classified by sales (sales classification) and strata classified by number of employees (employee classification). For all six divisions by industry and scale, the error ratio for sales decreases. By dividing the population into strata by sales, we can make a perfect ordering of groups by sales and thereby decrease the population variance of sales, particularly for real estate and some other sectors. Table 3 also includes the error ratios for fixed investment. The error ratios based on the sales classification are lower than those based on the employee classification.

This is because correlations between sales and fixed investment are stronger than those between the numbers of employees and fixed investment. From this result, it is reasonable to suppose that sales is a better measure of the "scale" of enterprises than the number of employees, and thus a better basis for stratified sampling in Japanese business surveys.

With this improvement in the sampling methodology, the Bank can reduce the necessary number of sample enterprises. In regular revisions of the Tankan sample enterprises, the Bank continues to use current sample enterprises to maintain continuity in the time series data and add new sample enterprises to satisfy criteria pertaining to statistical accuracy and fitness between sample enterprises and population enterprises. In the March 2015 revision, the Bank added over 1,000 new sample enterprises. Using the sales classification, the Bank would only have needed to add a few hundred new sample enterprises to satisfy the relevant statistical criteria. Moreover, statistical accuracy of results will improve even with a smaller sample.

We next check sample bias. We can see from Table 4 that sample biases are also lower for the sales classification as a consequence of declines in the error ratios. The reduction of bias is particularly pronounced for nonmanufacturing enterprises. Biases are also lower for fixed investment for five of the six divisions by industry and scale.

Table 4: Sample Bias based on the Strata Classified by Sales and Number of Employees

Item	Classified by	Large		Medium-sized		Small	
		Manuf.	Nonmanuf.	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.
Sales	Employees	2.3%	11.3%	3.3%	19.8%	11.4%	26.4%
	Sales	1.8%	1.6%	1.2%	2.5%	10.0%	9.6%
Fixed Investment	Employees	3.5%	9.8%	6.2%	20.2%	9.0%	13.7%
	Sales	3.4%	8.4%	10.9%	10.8%	6.6%	9.0%

However, there is still a problem of overestimation due to the aforementioned tendency for firms with fewer employees to be underrepresented in the Tankan. There is little scope for improvement here: if the Bank were to attempt to force very small companies to participate in the Tankan survey, the near-100% response rate would almost certainly be jeopardized.

That said, to grasp actual economic conditions accurately, the Bank must still look to reduce statistical bias to the greatest extent possible. Adding enterprises with relatively small number of employees to each stratum to reduce sample biases is not reasonable from the perspective of survey cost, with the number of sample enterprises already deemed adequate from the perspective of statistical accuracy. Eliminating a significant number of enterprises would also be controversial in as much as the post-revision results of the judgment survey, such as the "business conditions", would end up differing significantly from the pre-revision results. Differences may be especially large at the industry or region level, thereby creating unsatisfactory discontinuities in the time series data. Continuing to use almost all current sample enterprises appears to be a principled approach in practice.

Taking the above issues into consideration, we try to eliminate just a few percent of enterprises, focusing on those that contribute the most to overestimation of the number of employees. We can see from Table 5 that the biases for sales and fixed investment decrease in parallel with the reduction in bias for the number of employees. Moreover, this degree of sample reduction should not cause the post-revision Tankan results to differ too drastically from the pre-revision results.

Table 5: Decreasing the Sample Bias by Eliminating Responding Enterprises

Item	Sample Size	Large		Medium-sized		Small	
		Manuf.	Nonmanuf.	Manuf.	Nonmanuf.	Manuf.	Nonmanuf.
Sales	11,126	1.8%	1.6%	1.2%	2.5%	10.0%	9.6%
	10,826	0.4%	0.6%	0.3%	1.6%	8.2%	8.7%
Fixed Investment	11,126	3.4%	8.4%	10.9%	10.8%	6.6%	9.0%
	10,826	3.2%	5.1%	9.1%	11.2%	3.9%	8.4%

#### 4. Concluding Remarks

Above we have outlined how the Tankan sampling design is improved by using sales data from the *Economic Census* to divide the population into appropriate strata. Using this information from the “*Economic Census*,” we find that the true error ratio is higher than the estimated value which the Tankan used prior to the March 2015 revision, and that the Tankan results overestimate the population values. By stratifying based on sales, we can decrease both the error ratio and the sample biases reflected in the Tankan results.

Although the problem of overestimation cannot be completely eliminated, we believe this approach strikes an appropriate balance between statistical accuracy and survey costs, thereby enabling a further improvement in the survey methodologies of the Tankan.

#### References

- Research and Statistics Department Bank of Japan(2004):“Sample Design and Sample Maintenance of TANKAN”, [online]< [http://www.boj.or.jp/en/statistics/outline/notice\\_2004/nttk14.htm/](http://www.boj.or.jp/en/statistics/outline/notice_2004/nttk14.htm/)>
- Research and Statistics Department Bank of Japan(2015):“Regular Revision of the Tankan Sample Enterprises”, [online]< [http://www.boj.or.jp/en/statistics/outline/notice\\_2015/not150302a.htm/](http://www.boj.or.jp/en/statistics/outline/notice_2015/not150302a.htm/)>
- Shimizu, M., and M. Suga (2013): “Economic Statistics – Statistical Study for Industrial Activities and Price fluctuation” (in Japanese), Baifukan