Not to be released before 4:00 p.m.
on Monday, June 7, 2004

June 7, 2004
Research and Statistics Department
Bank of Japan

Sample Design and Sample Maintenance of TANKAN

I. Introduction

TANKAN, the Short-term Economic Survey of Enterprises in Japan, was revised in the March 2004 Survey (released on April 1, 2004). In addition to the regular revision of sample enterprises which is conducted every 5 years, the revision includes some changes in the framework of the survey.[1]

This paper introduces the sample design and sample maintenance of TANKAN in detail among the revision.

II. Sample Design

(1) Philosophy of Sample Design

TANKAN aims at grasping the current conditions of the domestic economy by a quarterly business survey which adopts the sample survey framework. Although a complete survey may be more precise than a sample survey, it is impractical to survey all enterprises in Japan every quarter. Consequently, TANKAN fixes a population[2] at a certain point of time and selects sample enterprises out of the population according to a certain statistical criteria.

---

[1] See "Revision of TANKAN in the March 2004 Survey - Comparison between the pre- and post-revision in the December 2003 Survey" (released on March 8, 2004) for details. As for the general explanation and FAQ of TANKAN, see the Bank of Japan's website (http://www.boj.or.jp/en/stat/tk/tk.htm).

[2] The population is approximately 220,000 private enterprises (excluding financial institutions) in Japan with at least 20 million yen in capital, based on the Ministry of Public Management, Home Affairs, Posts and Telecommunications' "*Establishment and Enterprise Census of Japan*" (conducted as of October 2001).

Since the results of the sample survey are estimated from samples, the survey has errors (sampling errors) by nature. Therefore, an appropriate sampling method (sample design) is required for the sample survey to reduce these errors.

We basically took over the methodology used for the previous revision in March 1999,[3] and revised it somewhat to enhance the statistical accuracy and decrease the burden on respondents. Some criteria are newly set for further accuracy. The outline of the sample design is as follows.

i) Setting an acceptable error range (a statistical accuracy target).

ii) Dividing the population into strata (for population estimation) by industry and scale to achieve the statistical accuracy target with less samples.

iii) Maintaining the current surveyed enterprises as sample enterprises, and thereafter adding new sample enterprises to the appropriate strata to meet the statistical accuracy target (Stratified Sampling).

iv) In every stratum, testing the fitness between the population distribution and sample distribution (and adding new enterprises if the fitness is rejected by the test).

Details of the sample design are as follows.

(2) Setting a Statistical Accuracy Target

To set a statistical accuracy target, we first introduce how TANKAN aggregates responses.

There are two types of survey items in TANKAN, "Judgement Items" and "Quantitative Items." "Judgement items" are qualitative ones which inquire about each enterprise's business conditions for example. In these items, responding enterprises choose one out of three choices. On the other hand, "Quantitative Items" ask such figures as Sales, Fixed Investment and Number of Employees.

[3] As for the previous revision in March 1999, see "The Methodology of the Sampling and Aggregation of the *ALL Enterprises Tankan*" (released on May 31, 1999).

As for the "Judgement Items," responses are simply aggregated. To be concrete, DI (Diffusion Index) is calculated using the following formula.

$$DI = \begin{pmatrix} Percentage\ share\ of\ \ enterprises \\ responding\ Choice\ 1 \end{pmatrix} - \begin{pmatrix} Percentage\ share\ of\ \ enterprises \\ responding\ Choice\ 3 \end{pmatrix}$$

Meanwhile, in the "Quantitative Items," the population is estimated, namely, the aggregates of the population are estimated by multiplying the mean of sample enterprise responses by the number of population enterprises (population estimation method). In practice, to calculate figures by industry (30 industries) and scale (three scales: Large, Medium-sized and Small), we divide the population into strata by industry and scale and calculate the aggregate of each stratum. The aggregates of the overall population are obtained by adding up the aggregates of all strata. This can be expressed in the following formula.

$$\begin{pmatrix} Population\ estimate \\ of\ \ each\ stratum \end{pmatrix} = \begin{pmatrix} Mean\ of \\ sample\ enterprise\ responses \end{pmatrix} \times \begin{pmatrix} Number\ of\ \ population \\ in\ stratum \end{pmatrix}$$

$$= \frac{\begin{pmatrix} Total\ of \\ sample\ enterprise\ responses \end{pmatrix}}{\begin{pmatrix} Number\ of \\ sample\ enterprise\ responses \end{pmatrix}} \times \begin{pmatrix} Number\ of\ \ population \\ in\ stratum \end{pmatrix}$$

$$\begin{pmatrix} Overall\ \ population \\ estimate \end{pmatrix} = \sum (Population\ estimate\ of\ \ each\ stratum)$$

As indicated above, while "Quantitative Items" estimate the population from sample data, "Judgement Items" simply aggregate the sample data without using the population estimation method.

These aggregation methods imply that the statistical accuracy target should be defined by using "Quantitative Items" rather than "Judgement Items," since sampling errors occur through the process of estimating the population as mentioned in (1).

Based on the reason above, we set the target on "error ratio" of population-estimated Sales (See Appendix 1 for the definition of error ratio). Although we already had the same target in the previous revision, we set more detailed targets on error ratio this time. TANKAN has set the targets on error ratio of 6 categories by industry (Manufacturing and Nonmanufacturing) and scale (Large, Medium-sized and Small) (3 % for Manufacturing and 5% for Nonmanufacturing[4]). TANKAN also sets nonbinding targets on error ratio by each industry (30 industries) and scale (Large, Medium-sized and Small) in this revision (approximately 10% for each industry by scale). These new targets are intended to further improve the statistical accuracy.

Statistical Accuracy Target (Upper limit of error ratio of Sales<population estimate>)

	Large Enterprises	Medium-sized Enterprises	Small Enterprises
Manufacturing	3%	3%	3%
Nonmanufacturing	5%	5%	5%
Industry by industry (Nonbinding)	approximately 10%	approximately 10%	approximately 10%

—— Among major survey items in TANKAN, Sales is used for calculating error ratio, because approximate data for overall population enterprises are available. In order to calculate error ratio, data of sales corresponding to overall population are needed. In TANKAN, the population mean and population variance of Sales for each stratum are calculated by using the "Basic Survey of Japanese Business Structure and Activities" or "Census of Commerce" (both released by the Ministry of Economy, Trade and Industry).[5] Since both surveys are complete surveys, they can be used as substitute data for population information of TANKAN. Moreover, the correlation between Sales and Fixed investment is generally high (Figure 1). As Fixed investment is one of the most popular items for TANKAN users, it suggests that a sample design with Sales is useful.

[4] As individual enterprises vary largely overall in the nonmanufacturing compared to the manufacturing, the accuracy target is set less strictly.
[5] As with industries excluded from both surveys, the unbiased estimators of the population mean and population variance are obtained by using Sales of TANKAN sample enterprises.

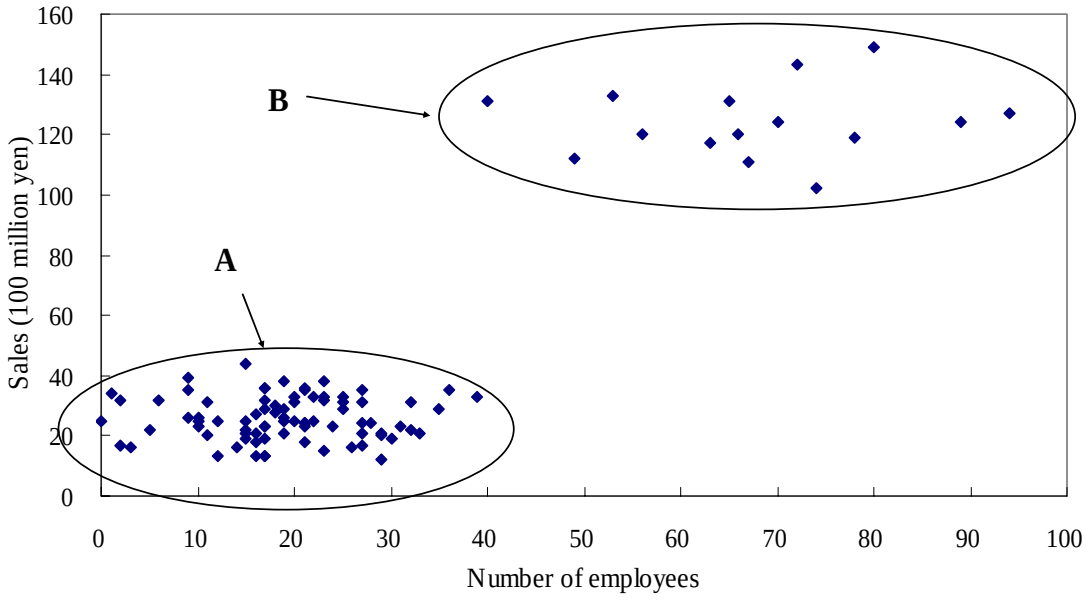
(3) Division of Stratum, Stratified Sampling

In order to attain the accuracy target with a less number of samples, it is more effective to further divide the population (already divided by industry and scale) into smaller strata, in other words, categorize enterprises into strata by characteristics.

In general, a smaller variance of each stratum can lower error ratio and decrease the number of sample enterprises. Therefore, if population is further divided to decrease the variance of each stratum, the accuracy target can be achieved with a less number of sample enterprises.

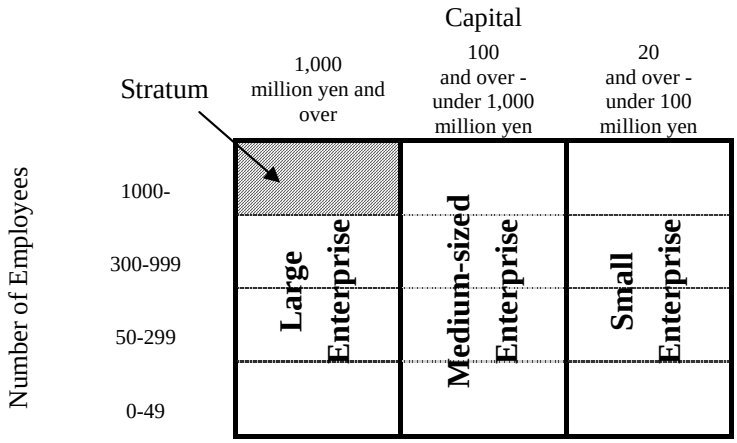
——For instance, as shown in the next chart, let us assume that sales of a certain industry are divided into two groups according to the number of employees. If we regard all enterprises in the industry as one stratum, the Sales variance of the stratum is large, because the Sales mean of A and B are quite different. For this reason, we have to select many samples to meet the accuracy target of the industry. On the contrary, assume that A and B are different strata, and then the Sales variance of each stratum, A and B, becomes small. Consequently, the accuracy target can be met with a less number of sample enterprises.

Image: Relationship between Employees and Sales in an Industry



According to the reason above, the pre-revision stratum has been classified by the industry and number of employees. After the revision, enterprises' capital has been added, making it to three bases in total (Figure 2 and 3). As a result, the number of strata was increased from 118 prior to the revision to 377.

—— In principle, population of each industry is divided into 3 strata for capital (20 -100 million yen, 100 million -1,000 million yen, 1,000 million yen and over) by 4 strata for number of employees (0-49 employees, 50-299 employees, 300-999 employees, 1,000 employees and over) (So each industry is divided into 3*4=12 strata). For some industries, the population is divided further.[6] The relationship of these strata to scale (Large, Medium-sized and Small) can be described in the chart below (Each cell corresponds to each stratum).



—— For example, when looking at the average Sales by stratum in Transportation, Sales differ substantially according to the number of employees, even though the size of capital is the same. This shows that it is effective to divide the strata by using both the capital and number of employees (Figure 4).

Because the population is further divided compared to the pre-revision TANKAN, there have been strata in which there are only a few population enterprises. For instance, strata with very large capital and a very small number of employees can be in such a situation.

For these strata, there are possibilities that no response is obtained and population

[6] For example, the stratum of Large Electrical machinery is divided into two strata by capital scale (Figure 2).

estimate can not be calculated, even though we take in all the enterprises in the stratum as sample.

To avoid such possibilities, we basically do not extract any samples from 40 strata with less than five population enterprises.[7,8]

Now let us extract sample enterprises from each stratum (337 strata, excluding the above 40 strata). Following the previous revision, firstly we maintain the current surveyed enterprises as sample enterprises,[9] thereafter we add new sample enterprises at random from the appropriate strata to meet the accuracy target.

—— Although theoretically sampling all enterprises at random in every revision is ideal, practically it is difficult to implement it, because we have to add a very large number of new sample enterprises. Moreover, from another point of view, there are some merits to maintain the current sample enterprises such as keeping a high response rate and avoiding wrong responses.

In addition to the targets above, we also set a nonbinding target for sampling ratio (= (Number of sample)/(Number of population)) in this revision. The sampling ratio of each stratum is targeted at approximately one percent or over.[10] This target aims at avoiding disturbance caused by population estimation process. Population estimation aggregates the result using sampling ratio, namely, the aggregate is calculated as (Total of Sample responses)/(Sampling ratio[11]). For example, if the sampling ratio is 5 percent in a stratum, figure of each enterprise is inflated 20 times when we aggregate population estimate. Consequently, if sampling ratio is too low, it may cause a disturbance to the aggregate due to the fluctuation of some enterprises.

---

[7] Strata with no population enterprise are excluded from the calculation of error ratio, and these 40 strata are also excluded from the calculation.

[8] As for these 40 strata, we adopt an imputation, which substitute the mean of a neighboring stratum to those of 40 strata. There are also possibilities that no response is obtained from a stratum, even if the stratum has 5 or more population enterprises. In such case, we adopt the same imputation.

[9] Except for enterprises whose capital is under 20 million yen. These enterprises are excluded from the population of the survey in this revision.

[10] We will keep sampling ratios at 0.5 percent or over, even if the number of samples would decrease.

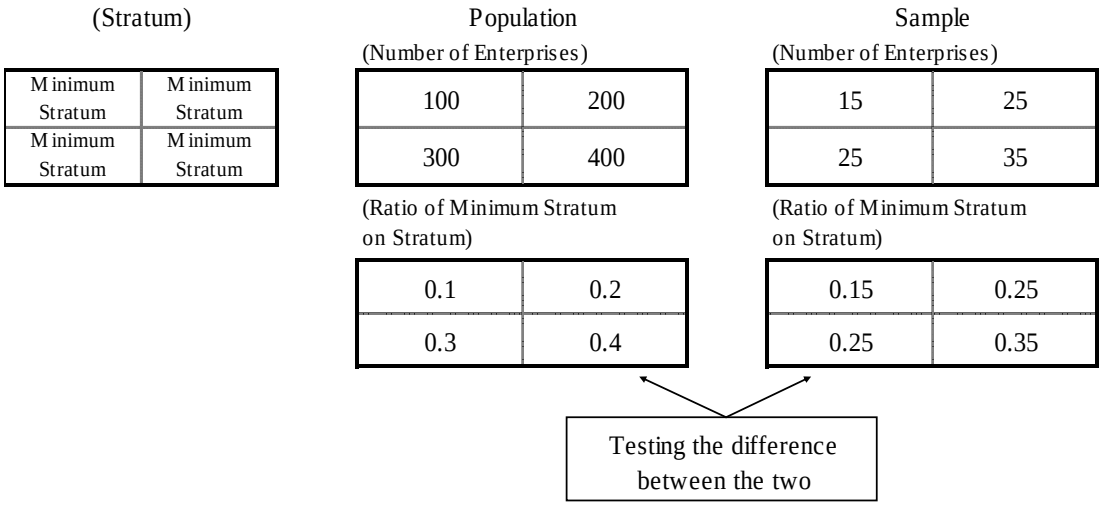
[11] Exactly, the ratio is (Number of respondents out of samples)/(Number of population).

(4) Testing Fitness between Population and Sample Distributions

The population estimates of TANKAN may possibly be biased since not all samples are extracted at random. To solve the problem, TANKAN tests the fitness between the population and sample distributions. If sample distribution is significantly unfit, we add new samples in the same way as in the previous revision to prevent the biased estimation.

To be concrete, each stratum is further divided into several "Minimum strata" by enterprises' capital and number of employees. After that, the difference between the population and sample distributions is tested by the chi-square goodness of fit test (Appendix 2). When there is a significant bias within a stratum, samples are added to adjust the distribution. This examination has enabled us to estimate the population with hardly any bias.

Image: Fitness Test of Population and Sample Distribution



(5) Results of Sample Design

Following the above sample design, the sample enterprises in the March 2004 Survey, the first survey after the revision, are 10,562 (Figure 5). Among all sample enterprises, 6,964 are continuing samples and 3,598 are added in this revision.

Chronology -- Number of Sample Enterprises

Revision Date	Number of Sample
May 1974 (Start date of TANKAN)	5,596
May 1980 (1st revision)	5,373
May 1983 (2nd revision)	7,035
February 1985 (3rd revision)	7,116
May 1988 (4th revision)	7,635
Nov 1993 (5th revision)	10,011
March 1999 (6th revision)	9,433
March 2004 (Current revision)	10,562

In terms of the statistical accuracy, all the binding targets have been achieved and new nonbinding targets have been also achieved by and large.[12]

- Error Ratio of the 6 Categories (binding)

The targets have been totally achieved.

Error Ratio of Sales (Population Estimate)

	Large Enterprises	Medium-sized Enterprises	Small Enterprises
Manufacturing (target: 3% or less)	1.2%	2.7%	2.7%
Nonmanufacturing (target:5% or less)	2.4%	3.6%	3.1%

- Error Ratio for each industry: approximately 10% or less (nonbinding)

Most categories by industry and scale have met the accuracy target except for Mining (3 categories out of 90). Another 13 strata have also not met the target, but error ratio is under 12% (Figure 5).

[12] There is a trade-off between achieving nonbinding target (error ratio for each industry and sampling ratio for each stratum) and decreasing the burden on respondents. In this revision, we have not completely achieved these nonbinding targets, since it turned out that quite a few samples are required to achieve the targets.

- Sampling ratio for each stratum: 1% or over (nonbinding)

Most strata (331 strata out of 337) have achieved the target. The other 6 strata have not reached the ratio of 1%, but have reached to around 0.9%.

- Fitness of the distribution (binding)

The targets have been totally achieved in all 337 strata.

(6) Sample Design for Financial Institutions

For financial institutions, we introduced the sample survey framework in this revision and have started to survey "Shinkin banks," "Other financial institutions for small businesses," and "Non-deposit money corporations" in addition to the current surveyed categories: "Banks," "Securities companies," and "Insurance companies." Details of the sample design are as follows.

—— The population of the survey consists of 7 business categories: i) "City banks, Long-term credit banks & Trust banks[13]," ii) "Regional banks & Regional banks II," iii) "Shinkin banks," iv) "Other financial institutions for small businesses," v) "Securities companies," vi) "Insurance companies," and vii) "Non-deposit money corporations[14]." The total population consists of 808 enterprises.[15]

—— 10 percent or less error ratio of Fixed investment[16] (population estimate) for overall financial institutions is set as a binding target. There is no target for each business category.

—— For each business category, the population is divided into strata by asset[17] to decrease the variance of each stratum. There are 18 strata in total (Figure 6). Note that population is not divided into three scales, Large, Medium-sized and Small.

[13] Includes domestic trust banks which started operations after October 1993 and foreign capital ones, and new-type banks which specialize in financial intermediation using websites and ATMs and do not belong to traditional banking sectors.
[14] The population is the stock-listed enterprises.
[15] Holdings companies are not included.
[16] Sales are not surveyed for financial institutions, so we set the error ratio of Fixed investment instead.
[17] We adopt assets because it is more closely related to fixed investment.

—— Sample enterprises are extracted from each stratum to achieve the accuracy target (stratified sampling).

—— Fitness between the population and sample distributions is tested for each stratum by chi-square goodness of fit test. If there is a significant unfitness in a stratum, new samples are added to the stratum.

Following the above sample design, the number of sample enterprises in the March 2004 Survey was a total of 210 enterprises (Figure 5). Among all sample enterprises, 111 are continuing samples and 99 are added in this revision.

In terms of the statistical accuracy, all targets have been achieved. The error ratio of Fixed investment for overall financial institutions is 3.3 percent, lower than the target of 10 percent. As for the fitness of distributions, there was no significant unfitness for all strata.

—— For the published figures, the population is divided into not 7 business categories but into 5, "Banks (i and ii above)," "Shinkin banks & Other financial institutions for small businesses (iii and iv above)," "Securities companies," "Insurance companies," and "Non-deposit money corporations." This aims at preventing a significant discontinuity of the statistics even when the mergers and spin-offs between different business categories occur.

III. Sample Maintenance

As mentioned in the section II, TANKAN extracts sample enterprises from the population fixed at the base point (the population based on "Establishment and Enterprise Census of Japan," by the Ministry of Public Management, Home Affairs, Posts and Telecommunications).

Recently, the structure of the Japanese economy and industry has been changing rapidly. In order for TANKAN to capture the actual business conditions, we will shorten the interval between the revisions of sample enterprises from every 5 years to every 2 or 3 years. [18,19]

[18] It has become possible because "Establishment and Enterprise Census of Japan" increased its frequency from once in every 5 years to twice. The next census is conducted as of June 1, 2004.
[19] Population and sample enterprises for financial institutions will also be revised every 2 or 3 years.

Even after the interval of revisions is shortened, both population and sample enterprises change during an interval between two revisions, due to various factors such as bankruptcies, newly established enterprises, and mergers and spin-offs. However, it is impossible for TANKAN to keep track of all population enterprises in detail, so TANKAN makes it a rule to fix the population at the revision time and maintains the sample enterprises during the interval between two revisions. Details of sample maintenance are below.

(1) Adding Sample Enterprises

The statistical accuracy may decline due to a decrease in the number of sample enterprises caused by bankruptcies, mergers and spin-offs, etc. To prevent lowering of statistical accuracy, we regularly (basically annually) check the statistical accuracy, such as error ratio and fitness of the population and sample distributions. When the accuracy is not enough to meet the targets, we add new sample enterprises.[20]

(2) Dealing with Mergers and Spin-offs

Business reorganizations such as mergers and spin-offs have been increasing. We basically take the following procedures to deal with mergers and spin-offs involving sample enterprises in TANKAN.

Mergers:

Only if the core enterprise (with most capital among the enterprises before the merger) is a sample, the enterprise after the merger is chosen as a sample.

- When there is a merger among the sample enterprises, or when a sample enterprise merges with a non-sample enterprise with less capital, the enterprise after the merger becomes a sample enterprise.
- When a sample enterprise merges with a non-sample enterprise with more capital, the enterprise after the merger does not become a sample enterprise.

[20] We will take the same procedures for financial institutions.

Spin-offs:

Only the core enterprise (with the most capital among the enterprises after the spin-offs) is chosen as a sample. - As an exception, if there are other enterprises within the same stratum as the core enterprise and the discontinuity (in population estimate of Sales or Fixed investment) becomes smaller when we choose them as samples, these enterprises are chosen as samples in addition to the core one.

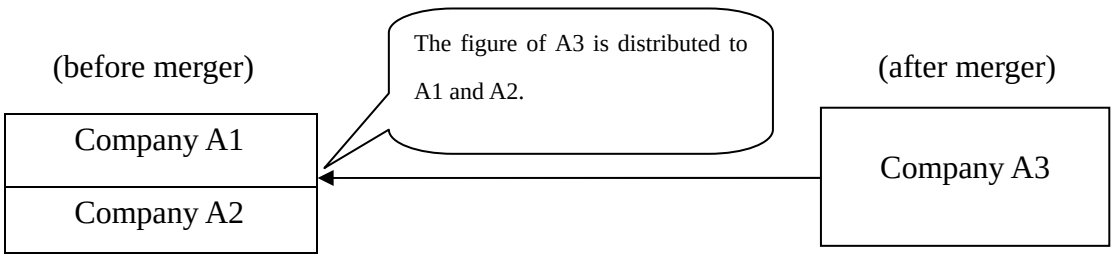
In this revision, we also introduce two exceptional rules to deal with mergers and spin-offs which have a certain or more impact on population estimate. These are: i) Distribution method and ii) Incorporating spin-off companies in the course of time.

i) Distribution Method

"Distribution method" in this paper is the method to distribute figures of enterprises after merger (or spin-off) into figures of the enterprises before merger (or spin-off). In other words, this method aggregates figures as if there are no mergers and spin-offs in order to decrease the technical fluctuations which do not reflect the actual conditions of the economy.[21] TANKAN adopts the method for mergers and spin-offs which have an impact on population estimate of Sales or Fixed investment to a certain degree. An image of Distribution method is as follows.

For mergers:

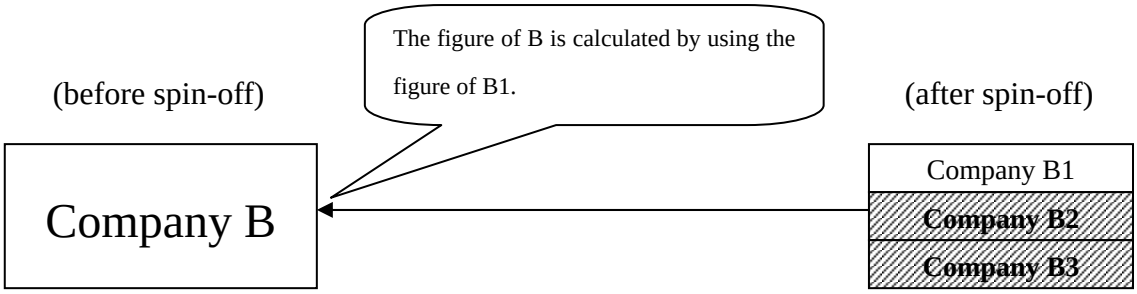
If companies A1 and A2 have merged together to be a company A3, first we ask for the figure of A3, and then distribute the figure into the figures of A1 and A2. To be concrete, the figure of A1 (A2) is calculated as the figure of A3 multiplied by the ratio, which is the figure for A1 (A2) divided by the figure for A3 at the time of the merger.



[21] Due to this process, the number of sample enterprises or responses in released materials may be different from the number of actual sample enterprises or responses after the merger or spin-off.

For spin-offs:

If a company B is spun off into B1, B2 and B3, we basically choose only the core enterprise (let us assume B1 is the one). In this case, first we ask for the figure of B1, and then calculate the figure for company B. To be concrete, the figure of B is calculated as the figure of B1 multiplied by the ratio, which is the figure of B divided by the figure of B1 at the time of the spin-off.



Among all mergers and spin-offs, we consider to apply this method only if the impact is large, namely, the impact on the population estimate (of Sales or Fixed investment) of the 6 categories (Manufacturing and Nonmanufacturing by Large, Medium-sized and Small) is approximately 1 percent or over. The category which we estimate the impact is the one which the enterprise (the core one in the case of mergers) prior to the merger or spin-off belongs to.

—— Spin-off with establishing a holding company is an example which may have a large impact on population estimates. The pure holding company, which does not have any actual business activity, may have the largest capital. As we choose only the core enterprise (which has the largest capital among enterprises after the spin-off) as a sample, the pure holding company may be often the core enterprise. In such case, the figures (Sales or Fixed investment) of the pure holding company are often much smaller than those of the pre-existing company, so it may have a large impact on population estimate.

—— As mentioned in the beginning of III, the number of population enterprises remains unchanged from the time of sample design until the next update of the population information. Based on this idea, it is natural to transform recent enterprises' figures into the figures of the samples at the time of sample design.

ii) Incorporating spin-off companies in the course of time

Even if an enterprise is not chosen as a sample enterprise at the time of the spin-off, we check whether the enterprise has grown and had great influences on the aggregates. If

such an enterprise after the spin-off satisfies a certain criteria, it is flexibly incorporated into sample enterprises, even if some time has passed.

To be concrete, we consider to incorporate the spin-off enterprise when the year-on-year figures of Sales or Fixed investment of the enterprise is larger than approximately 1 percent of population estimate in one category out of 6 (Manufacturing and Nonmanufacturing by Large, Medium-sized and Small) to which the core company belongs. We may incorporate such spin-off companies at the time of regular (basically annual) sample addition.[22]

—— For example, a non-core enterprise, which was not chosen as a sample at the time of spin-off, may increase fixed investment since its business has taken off. In such case, we can not capture the whole picture of the pre-existing sample enterprise any more if we do not take in the enterprise as sample.

(3) Dealing with Non-sampling Errors

It is important to reduce not only "sampling errors" but also "non-sampling errors" through sample maintenance. Non-sampling errors occur since the information obtained from samples is imperfect.[23] To reduce non-sampling errors, TANKAN is making efforts to increase the response rate and obtain accurate response.[24]

—— In practice, we are asking for the understanding and cooperation of sample enterprises to increase the response rate.[25] Moreover, after receiving responses from enterprises, contents of the responses are checked and confirmed carefully through interviews to prevent clerical errors and misunderstandings of the definition.

When the response of an enterprise is missing in spite of our efforts, we use missing-value supplement method, which substitutes a certain figure for a missing value as a rule. We changed the method in this revision.

—— Until this revision, unanswered responses (missing value) have been treated

[22] The number of population enterprises is changed exceptionally in this case.
[23] Non-sampling errors also exist in complete surveys.
[24] As sampling errors in TANKAN are calculated based on 100 percent response rate, increasing the response rate is also important in terms of reducing sampling errors.
[25] 96.5 percent (March 2004 Survey)

simply: the responses were excluded from the aggregation. For the quantitative items, which adopt the population estimation method, this treatment corresponds to substituting the mean of the stratum to the missing value[26]. However, it induces a technical fluctuation to the population estimate when the last response of the enterprise is far from the mean of the stratum.

Although the above method has not induced serious problems in TANKAN because of its high response rates, we have statistically examined alternatives to intend TANKAN's accuracy in the future. As a result, it turned out that substituting the most recent response of the corresponding enterprises to the missing value is generally better than substituting the mean of the stratum[27].

Therefore, we have introduced the new method to substitute the most recent response of the corresponding enterprises to the missing value for all annual projection items such as Sales, Current Profits and Fixed Investment[28].

[26] This is because the population estimation method calculates the aggregated figure by multiplying the mean by the number of population.

[27] The examination method used here is the same as the one used in "Methodology for Handling Missing Values In TANKAN" by Kiyohito Utsunomiya and Katsurako Sonoda (2001), Bank of Japan Research and Statistics Department Working Paper 01-11.

[28] The judgement items and other quantitative items (quarterly items and "Number of new graduates hired") are treated in the same way as the pre-revision method.

Image: Missing Value Supplement

Case: FY2004 Sales of Company A was missing at the September 2004 survey:

Pre-revision

<Sales of Company A>	FY 2003	FY 2004
June 2004 Survey	100	110
September 2004 Survey	100	**200**

Substitute the mean of the population estimation stratum to which Company A belongs for the missing value.

Post-revision

<Sales of Company A>	FY 2003	FY 2004
June 2004 Survey	100	110
September 2004 Survey	100	**110**

Substitute the most recent response of Company A for the missing value.

IV. Conclusion

It is important to provide accurate statistics for TANKAN users, and sample design and maintenance described in this paper are the basis for making accurate statistics.

For TANKAN, there are issues to deal with. For example, Japan's economic structure is changing rapidly, and business reorganizations such as mergers and spin-offs are increasing.

To deal with these issues, we have shortened the interval between sample revisions and introduced new rules for mergers and spin-offs in this revision.

We will continue to make efforts to enhance the accuracy of TANKAN from a wide perspective including sample design and maintenance.

For further information, please contact:
Economic Statistics Division
Research and Statistics Department
Bank of Japan
Email:post.rsd5@boj.or.jp

Appendix 1: Definition of error ratio

$$Error\ ratio \equiv \frac{Standard\ deviation\ of\ Sample\ mean}{Population\quad mean}$$

$$Standard\ deviation\ of\ Sample\ mean \equiv \sqrt{\sum_{i=1}^{L} W_i^{\,2}\,\frac{N_i - n_i}{N_i - 1}\,\frac{\sigma_i^2}{n_i}}$$

$$Population\ mean \equiv \sum_{i=1}^{L} W_i\ \overline{Y_i}$$

$$\begin{cases} N_i : number\ of\ \ population\ enterprises\ in\ stratum\ i \\ n_i : number\ of\ \ sample\ enterprises\ in\ stratum\ i \\ \overline{Y_i} : population\ mean\ (of\ \ Sales)\ in\ stratum\ i \\ \sigma_i^2 : population\ variance\ (of\ \ Sales)\ in\ stratum\ i \\ W_i : ratio\ of\ \ N_i\ to\ number\ of\ \ population \quad (*) \\ L : number\ of\ \ strata\ in\ \ population \quad (*) \end{cases}$$

*: When you calculate error ratio in each industry or scale, use "population in the industry or scale" instead of "population."

Appendix 2: Chi-square goodness of fit test

Chi-square goodness of fit test is a method which tests whether the difference between the two distributions is significant. In TANKAN, we compare two distributions; number of population enterprises and number of sample enterprises for each stratum. We test the null hypothesis: "the population distribution and the sample distribution are the same." The detailed procedure is presented below.

(1) Subdivide the stratum into several Minimum strata ($i = 1, 2, \cdots, j$) based on the capital and number of employees.

(2) Let $N_1, N_2, \cdots, N_j$ ($n_1, n_2, \cdots, n_j$) be the number of population (sample) enterprises in Minimum strata: $i = 1, 2, \cdots, j$.

(3) Calculate the composition ratio of population enterprises in each Minimum stratum:

$p_i \equiv \frac{N_i}{N} \quad (N \equiv \sum_{k=1}^{j} N_k)$.

(4) Under the null hypothesis, "the population distribution and the sample distribution are the same," the number of sample enterprises in each Minimum stratum is expected to be $e_i \equiv n \cdot p_i \quad (n \equiv \sum_{k=1}^{j} n_k)$. e_i is the expected frequency for the Minimum stratum i, while n_i is the observed frequency for i.

(5) Test the null hypothesis, "the population distribution and the sample distribution are the same," by chi-square goodness of fit test with one-side 5 percent critical region.

Figure 1

Correlation between Sales and Fixed Investment -- Classified by Industry (All Scales)

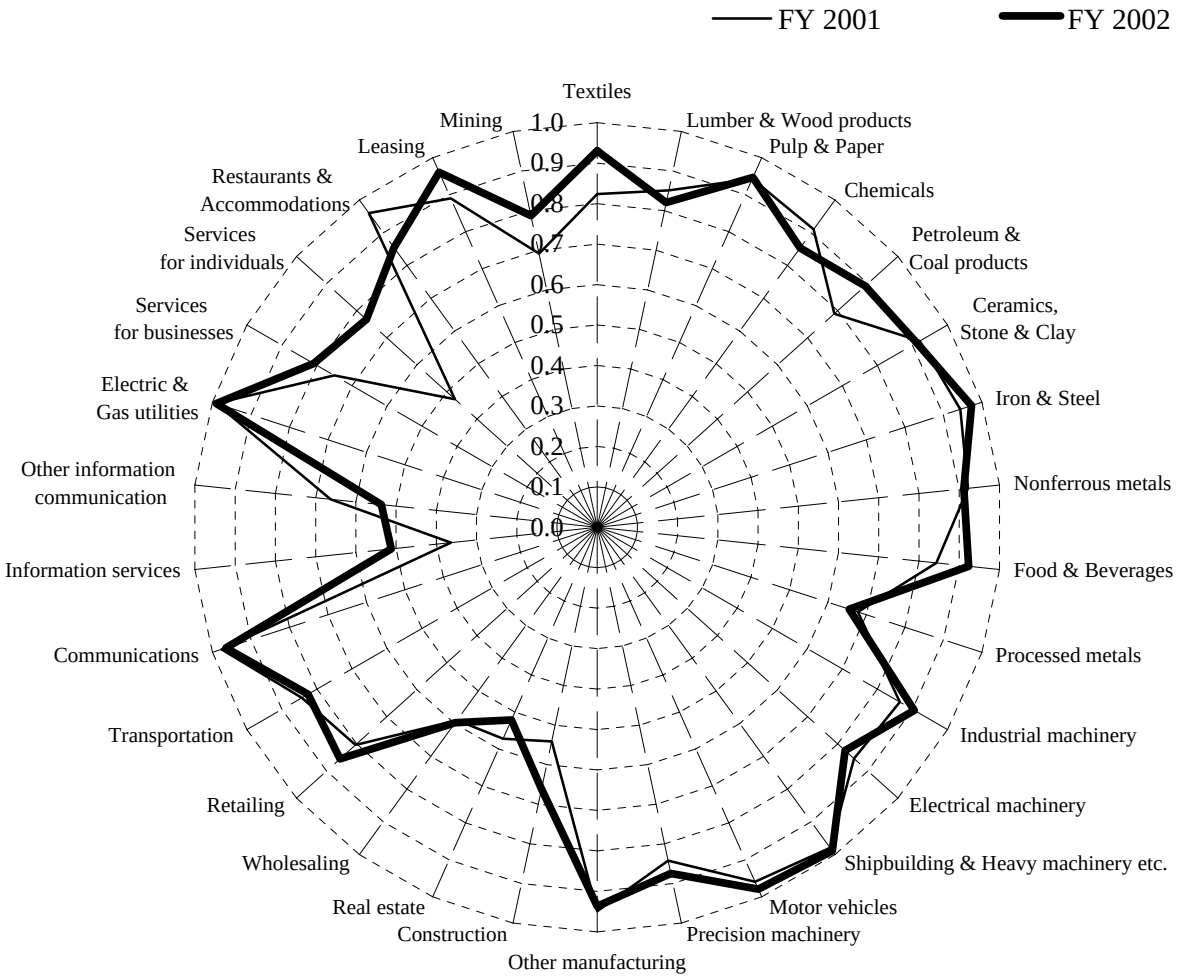
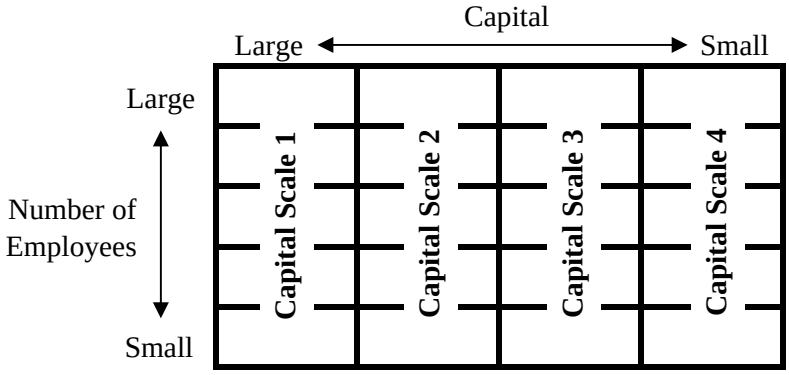


Figure 2

Stratum Division -- by Capital Scale

(1) Image (Each cell corresponds to each stratum.)



(2) Minimum Capital for each stratum

(Million yen)

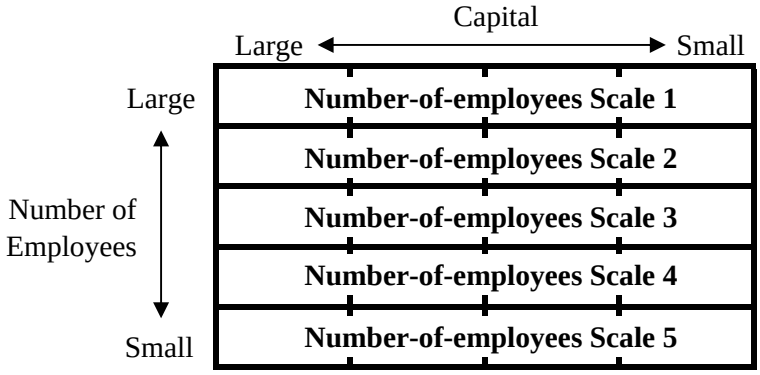
	Large enterprises		Medium-sized enterprises		Small enterprises	
stratum division	1	2	2	3	3	4
(Industry)						
Textiles	1,000		100		20	
Lumber & Wood products	1,000		100		20	
Pulp & Paper	1,000		100		20	
Chemicals	1,000		100		20	
Petroleum & Coal products	1,000		100		20	
Ceramics, Stone & Clay	1,000		100		20	
Iron & Steel	1,000		100		20	
Nonferrous metals	1,000		100		20	
Food & Beverages	1,000		100		20	
Processed metals	1,000		100		20	
Industrial machinery	1,000		100		20	
Electrical machinery	**100,000**	1,000		100		20
Shipbuilding, Heavy machinery & Other transportation machinery	1,000		100		20	
Motor vehicles	**50,000**	1,000		100		20
Precision machinery	1,000		100		20	
Other manufacturing	1,000		100		20	
Construction	1,000		100		20	
Real estate	1,000		100		20	
Wholesaling	**100,000**	1,000		100		20
Retailing	**20,000**	1,000		100		20
Transportation	1,000		100		20	
Communications	1,000		100		20	
Information services	1,000		100		20	
Other information communication	1,000		100		20	
Electric & Gas utilities	1,000		100		20	
Services for businesses	**15,000**	1,000		100		20
Services for individuals	1,000		100		20	
Restaurants & Accommodations	1,000		100		20	
Leasing	1,000		100		20	
Mining	1,000		100		20	

(*) Underlines indicate subdivided strata.

Figure 3

Stratum Division -- by Number of Employees

(1) Image (Each cell corresponds to each stratum.)



(2) Minimum Number of Employees for each stratum

(Number of Employees)

stratum division	Large enterprises	Medium-sized enterprises		Small enterprises		Others	
	1	2	3	3	4	4	5
(Industry)							
Textiles	1,000	300		50		0	
Lumber & Wood products	1,000	300		50		0	
Pulp & Paper	1,000	300		50		0	
Chemicals	1,000	300		50		0	
Petroleum & Coal products	1,000	300		50		0	
Ceramics, Stone & Clay	1,000	300		50		0	
Iron & Steel	1,000	300		50		0	
Nonferrous metals	1,000	300		50		0	
Food & Beverages	1,000	300		50		0	
Processed metals	1,000	300		50		0	
Industrial machinery	1,000	300		50		0	
Electrical machinery	1,000	300		**100**	50		0
Shipbuilding, Heavy machinery & Other transportation machinery	1,000	300		50		0	
Motor vehicles	1,000	300		50		0	
Precision machinery	1,000	300		50		0	
Other manufacturing	1,000	300		50		0	
Construction	1,000	300		50		**20**	0
Real estate	1,000	300		50		**15**	0
Wholesaling	1,000	100		**60**	20		0
Retailing	1,000	**200**	50		20		0
Transportation	1,000	300		50		0	
Communications	1,000	300		50		0	
Information services	1,000	**300**	50		20		0
Other information communication	1,000	**300**	50		20		0
Electric & Gas utilities	1,000	300		50		0	
Services for businesses	1,000	50		20		0	
Services for individuals	1,000	**100**	50		20		0
Restaurants & Accommodations	1,000	**300**	50		20		0
Leasing	1,000	**300**	50		20		0
Mining	1,000	300		50		0	

(*) Underlines indicate subdivided strata.

Figure 4

Relationship between Capital and Number of Employees and Sales -- Transportation

Average sales (Million yen)

		Capital		
		1,000 million yen and over	100 and over - under 1,000 million yen	20 and over - under 100 million yen
Number of Employees	1,000-	228,934	23,804	18,582
	300-999	41,069	9,670	5,605
	50-299	17,276	5,113	2,217
	0- 49	3,935	1,607	437

Source: Annual sales of TANKAN sample enterprises for fiscal 2001.

Figure 5

Number of Population and Sample Enterprises and Error Ratio by Industry and Scale -- 2004 March Survey

(1) Enterprises

	Population	Sample
Total	215,252	10,562

Industry	Large Enterprises			Medium-sized Enterprises			Small Enterprises		
	Population	Sample	Error Ratio	Population	Sample	Error Ratio	Population	Sample	Error Ratio
Manufacturing	2,078	1,187	1.2%	5,802	1,163	2.7%	37,703	1,940	2.7%
Textiles	72	49	3.6%	253	62	8.5%	3,138	141	7.0%
Lumber & Wood products	30	13	6.3%	149	35	9.4%	1,949	89	8.7%
Pulp & Paper	46	32	5.1%	178	43	8.9%	1,115	82	9.5%
Chemicals	295	142	4.6%	481	63	8.4%	1,241	76	9.8%
Petroleum & Coal products	29	22	6.4%	36	22	11.0%	152	30	11.2%
Ceramics, Stone & Clay	92	49	3.5%	325	56	9.0%	2,819	122	9.8%
Iron & Steel	82	54	5.0%	240	49	8.5%	905	102	11.1%
Nonferrous metals	65	38	5.4%	151	56	8.4%	528	67	11.6%
Food & Beverages	206	119	6.9%	728	120	8.6%	5,040	202	7.7%
Processed metals	94	56	4.2%	368	72	6.4%	3,690	148	8.7%
Industrial machinery	270	149	5.9%	700	118	6.1%	4,941	221	7.1%
Electrical machinery	384	202	1.8%	917	192	7.6%	3,792	257	10.3%
Shipbuilding, Heavy machinery & Other transportation machinery	37	31	1.6%	98	48	7.0%	477	73	10.5%
Motor vehicles	157	117	1.6%	315	84	6.4%	1,245	88	6.9%
Precision machinery	62	39	6.0%	154	45	8.4%	920	93	9.6%
Other manufacturing	157	75	5.6%	709	98	7.6%	5,751	149	6.5%
Nonmanufacturing	3,394	1,253	2.4%	16,550	1,695	3.6%	149,725	3,324	3.1%
Construction	304	143	6.3%	1,619	203	4.5%	53,521	857	6.2%
Real estate	424	92	7.7%	1,911	161	10.7%	10,560	148	10.8%
Wholesaling	717	195	6.3%	4,225	386	9.2%	29,952	648	6.8%
Retailing	389	159	7.8%	1,866	216	9.1%	16,473	411	8.3%
Transportation	310	143	7.6%	1,260	164	6.5%	10,003	343	5.1%
Communications	96	42	9.0%	122	51	8.7%	260	38	11.6%
Information services	220	92	6.0%	1,349	79	9.8%	3,666	89	10.6%
Other information communication	246	54	8.2%	641	76	5.6%	1,495	73	8.9%
Electric & Gas utilities	82	46	0.6%	133	56	11.1%	109	37	11.3%
Services for businesses	132	73	8.2%	1,030	85	9.4%	8,641	166	9.8%
Services for individuals	176	86	4.2%	1,204	59	9.1%	8,234	206	9.8%
Restaurants & Accommodations	213	75	7.2%	970	77	9.0%	5,539	158	8.1%
Leasing	60	36	7.7%	198	70	9.2%	961	99	11.6%
Mining	25	17	22.1%	22	12	25.0%	311	51	23.8%

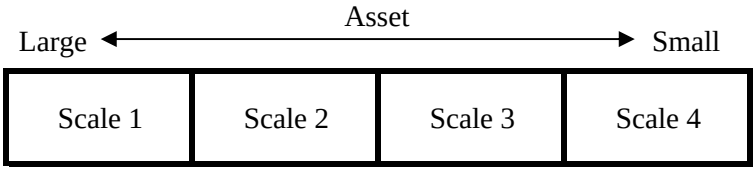
(2) Financial Institutions

	Population	Sample	Error Ratio
Total	808	210	3.3%
Banks	162	80	—
Shinkin banks & Other financial institutions for small businesses	321	35	—
Securities companies	227	27	—
Insurance companies	66	46	—
Non-deposit money corporations	32	22	—

Figure 6

Stratum Division for Financial Institution (by Asset Scale)

(1) Image (Each cell corresponds to each stratum.)



(2) Minimum Asset for each stratum

(Billion yen)

Scale	1	2	3	4
City banks, Long-term credit banks & Trust banks	40,000	5,000	0	—
Regional banks & Regional banks II	5,000	1,000	0	—
Shinkin banks	400	0	—	—
Other financial institutions for small businesses	0	—	—	—
Securities companies	1,000	50	0	—
Insurance companies	10,000	1,000	200	0
Non-deposit money corporations	1,000	0	—	—