Bank of Japan Working Paper Series

Treatment of Outliers in Business Surveys: The Case of *Short-term Economic Survey of Enterprises in Japan* (*Tankan*)

Atsushi Ishikawa*
atsushi.ishikawa@boj.or.jp

Shunsuke Endo†
shunsuke.endou@boj.or.jp

Tetsuya Shiratori‡
tetsuya.shiratori@boj.or.jp

No.10-E-8
September 2010

Bank of Japan
2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-8660

* Research and Statistics Department, Bank of Japan
† Bank of Japan
‡ Research and Statistics Department, Bank of Japan

Treatment of Outliers in Business Surveys:
The Case of *Short-term Economic Survey of Enterprises in Japan* (*Tankan*)

Atsushi Ishikawa*
Shunsuke Endo†
Tetsuya Shiratori‡

September 2010

[Abstract]

This paper discusses the appropriate treatment of outliers in *Tankan*. Under the current population estimate method, if one of the data which greatly diverges from the rest of the data (the so-called "outlier") does not represent the population, the estimates could also greatly diverge from the real value for the population. Since the real picture of the population when the outlier occurred is unknown, it is difficult to choose the optimum method through empirical analyses. Considering 1) the accuracy of *Tankan*, 2) the maintenance of its "easy to understand" quality, and 3) the reduction of its work load, we assume that outliers in *Tankan* are those which have extremely large impact on the rate of change from the previous year or the rate of revision from the previous survey of the estimated values for the population. The method of detecting and treating outliers is suitable for the characteristics of *Tankan* through empirical analyses of actual *Tankan* data. Specifically, detection method is based on the adjusted "range edit" method, which uses 1 percentile and 99 percentile, so that it can be applied to non-parametric data, for which the incidence of "0" (zero) is very high. The outliers thus detected are regarded as missing values and should be augmented by applying the missing value imputation, which is already being used for *Tankan*.

Keywords: TANKAN; outlier; stratified sampling; weight; stratum jumpers

---

We are grateful for helpful discussion and comments from Hiroshi Saigo, Hiroe Tsubaki, Hibiki Ichiue, Seisaku Kameda, Takatoshi Sekine, Kazuo Monma, as well as the research staff at the Bank of Japan. The views expressed in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Bank of Japan.

* Bank of Japan; E-mail: atsushi.ishikawa@boj.or.jp
† Bank of Japan; E-mail: shunsuke.endou@boj.or.jp
‡ Bank of Japan; E-mail: tetsuya.shiratori@boj.or.jp

1. Introduction

The content of *Short-term Economic Survey of Enterprises in Japan* (hereafter referred to as "*Tankan*") comprises judgment items, in which the diffusion indices of business conditions, etc. are obtained, and numerical items, such as sales and the amount of fixed investment. With respect to the latter, the estimates for the population are obtained based on values reported from sample enterprises which are chosen through stratified sampling. The estimated values for the population and their rates of increases from the previous year and the rates of revisions from the responses in the previous survey are calculated and released by the Bank of Japan. In doing so, even if a value reported by a sample enterprise diverges greatly from the values reported by other sample enterprises in the same stratum, it is included in the calculation without taking any special measures, unless it is an error.

Under this method, however, if one of the data which greatly diverges from the rest of the data (the so-called "outlier") does not represent the population, the estimate for the population obtained from the calculation could also greatly diverge from the real value for the population. Although the real value for the population is not known as the survey does not cover all members of the population, it is important to examine the appropriateness of inclusion of such data. If outliers which diverge greatly from the rest occur in the future, their influence could become un-negligible to the analysts. Therefore, this paper discusses the appropriate treatment of outliers.

The composition of this paper is as follows. Chapter 2 considers what constitute outliers and how to treat them in sample surveys, and reviews the cases of the applications of various approaches. Chapter 3 discusses the mechanisms of occurrences and characteristics of outliers in *Tankan* as well as the problems in applying to *Tankan* data the approaches to outliers discussed in Chapter 2. It also discusses concrete approaches to avoid such problems in treating outliers in *Tankan* surveys. Chapter 4 uses actual data from *Tankan* surveys to empirically analyze the approaches arrived at in Chapter 3. Chapter 5 sums up the paper.

2. Countermeasures Against Outliers and Precedents

(1) The Definition of An Outlier

An outlier is one of the data which diverges from those which comprise the majority of data, and which occurs in sample surveys. Sometimes a response from a sample diverges greatly from the data reported from other samples, at other times, the value weighted by the weight of the sample diverges from the data group [Fuller (2009)]. The weight here means the reciprocity of the probability of a member of the population chosen as a sample. It means to what extent a sample represents the units in the population (how many units) [Lohr (1999)].

Statistical treatment of outliers has been studied in the fields of quality control and medicine for some time, but it has also been studied in the field of business surveys, especially in other countries. Studies on outliers in business surveys have begun in full gear since the 1980s and Chambers (1986) has classified them systematically. According to the paper, outliers in sample surveys are either (1) representative or (2) non-representative. The former is a value which is "representative (one which cannot be regarded to be unique in the population)," while the latter is a value which is "non-representative (one which can be regarded to be unique in the population)." This classification has helped the researchers who came later. Since actual data could include erroneous responses, such data have been classified as non-representative outliers. Consequently, in working with data, the correctness of the reported values must be verified, and if they are errors, they must be corrected. If a reported value is correct but constitutes an outlier, its appropriate treatment needs to be considered. Needless to say, since the real picture of the population is unknown, it is difficult to determine whether an outlier is representative or non-representative. If information is available only on a sample enterprise, one cannot determine that the other enterprises in the population do not show values close to the outlier (i.e., the outlier is non-representative).

(2) The Causes of Outliers

As mentioned earlier, one of the causes of outliers in sample surveys is errors in writing out the response or misinterpretation of the question, which results in an erroneous response. These responses need to be corrected. Even when a correct response is given, it can be an outlier. The causes of outliers which are correct responses to the questions can be classified into two categories.

One is the method of choosing samples. In stratified sampling, for example, when stratification is not accurate, a response could diverge from the responses of other samples. Stratification is made based on the attributes, etc. of the units in the population. However, since the capacity for gathering information on the population is limited, stratification is always made on limited information. This makes it impossible to eliminate the possibility of units which are different from the other units in the same stratum straying into the group. The method of choosing samples must be studied carefully in order to prevent the occurrences of outliers of this kind.

The second cause is a large change in reported values due to a time lag between the time when original samples were drawn and the time when these samples are used in a following survey. Even if the stratification in the original sampling is accurate and reported values and weights then are at appropriate levels, reported values could change greatly in a subsequent survey due to the time lag, and reported values and weighted data could diverge from other data. For example, in a business survey, fast growth of an enterprise can greatly change its condition from that at the time of an earlier sampling. In stratified sampling, such an enterprise is considered to have jumped out of the original stratum because of its rapid changes and moved to another stratum. It is possible to say that this is the deterioration of a sample design due to aging, while the enterprise in question is called a "stratum jumper."

Here is an example of a stratum jumper. Take a company which is a sample in stratum h (weight: 50), the fixed investment amount of which is 100 million yen in fiscal year t. If its fixed investment amount in fiscal year t+1 is 10 billion yen, based on its level of fixed investment amount, it is considered to have changed to an enterprise which should be in stratum h' (weight 5). In this case, it should be weighted by the weight of the stratum h' (5), but in practice, it is weighted by the weight of stratum h (50). As a result, its fixed investment amount after weighting is 10 billion yen x 50 = 500 billion yen (whereas it should be 10 billion yen x 5 = 50 billion yen). Consequently, the margin of change from the previous fiscal year (after weighting) becomes 500 billion yen – 5 billion yen = 495 billion yen. This could result in great divergences from reality in the fixed investment amount of the whole stratum (estimated value for the population) and its rate of change from the previous fiscal year.

Thus, countermeasures against outliers which result from the passing of time, as well as those which occur at the time of sample drawing, are an important point in treating outliers [Pfeffermann and Rao (2009)].

(3) Consistency with the Objective of Outlier Treatment

Outliers strongly influence the estimates for the population as a whole. As such, the estimates for the population can change greatly depending on how outliers are treated. Therefore, the objective of outlier treatment must be clear, and the treatment method must be suitable for attaining this objective.

In sample surveys generally, samples are designed in such a way as to render the bias in the estimated value zero (to obtain an unbiased estimated value) by adopting, for example, simple random sampling (SRS). This means that the estimated values arrived at after outlier treatments lose their unbiasedness. On the other hand, it can curb the variance in estimated values.

While there is this tradeoff between bias and variance, the objective of treating outliers can be said to make the estimates for the population as close as possible to the real values for the population. A frequently adopted measurement for the appropriateness of estimates is mean square error (MSE).

In practice, however, since the real population values are not known in most sample surveys, it is impossible to calculate the MSE. This means that there is no single "right answer" to the treatment methods of outliers. It is necessary to conduct empirical analyses, simulations, etc., which are suitable to the attributes of the data, to arrive at the optimum treatment of an outlier.

(4) Treatment Methods for Outliers

The question in introducing outlier treatment is what standard to adopt in determining that one of the data is an outlier. When one of the data is judged to be an outlier, the problem then is to choose from numerous alternatives, which may include (1) use the one of the data as it is in the computation, (2) make it closer to other data, and (3) exclude it from the computation. Therefore, the countermeasures against outliers can be divided into (1) the detection of outliers and (2) their treatment. This classification will be used in the following discussion of countermeasures against outliers.

(A) Detection of Outliers

The detection of outliers is a quantitative judgment, which requires an indicator of the degree of divergence of each data. Various methods of computing such indicators have been developed (Table 1).

The most basic methodology is "range edit." Assuming that x_i denotes data for sample i, that m denotes the scale of the position (the mean, and other representative values of the frequency distribution), and that s denotes the scale of scattering (e.g. variance), then d_i (the indicator of divergence) can be shown by Equation (1).

$$d_i = \frac{x_i - m}{s} \tag{1}$$

When d_i exceeds the upper limit or lower limit set in advance, it is detected to be an outlier. This method is called "range edit." In other words, by setting a tolerable range ($m - C_L s$, $m + C_U s$) by using constants C_L and C_U, x_i is judged to be an outlier when it is outside of this range.

In "range edit," however, if a mean or a standard deviation which is not robust against the outlier is used as the scale of position (m) or the scale of scattering (s), the tolerable range can fluctuate widely (i.e., d_i fluctuates widely), resulting in non-detection of outliers which diverge slightly. This is called the "masking effect." As methods which are robust against the outlier and which can solve the problem of "masking effect," non-parametric methods adopting the scale of position (m) and the scale of scattering (s) which are robust against outliers, have been devised.

A representative example of such methods is "quartile method." Under this method, assuming that q_2 denotes the median, that q_1 denotes the first quartile, that q_3 denotes the third quartile, the scale of position is set as q_2, the scales of scattering as $(q_2 - q_1)$ and $(q_3 - q_2)$, and the tolerable range as $[q_2 - C_L (q_2 - q_1), q_2 + C_U (q_3 - q_2)]$. The quartile method is a non-parametric method robust against outliers and is a type of "range edit." In practice, as the method of setting C_L and C_U is important, it needs to be set through empirical analyses, etc. of past data [Bernier and Nobrega (1998)].

Another "range edit" method takes note of the ratio to the value reported in the previous survey. In a survey taken at period t, r_{it}, which is the ratio of the value reported by sample i in the present survey (period t) to the value reported by i in the previous survey (period t - 1), is used to detect outliers. For example, if the value reported by sample i in the present survey is very large compared with that reported by i in the previous survey, r_{it} will be very large, and the present reported value will be detected as an outlier.

The use of the ratios, however, requires caution. In business surveys generally, the smaller the enterprise size, the more likely for its r_{it} to fluctuate. For this reason, compared with large enterprises, only small enterprises tend to be detected as outliers. This tendency is called the "size masking effect." As the distribution of r_{it} is from zero (the left end) to theoretically infinite (the right end) (when the reported value is a non-negative number), it has been pointed out that it is difficult to detect outliers from the left end of the distribution.

Meanwhile, Hidiroglou and Berthelot (1986) have proposed a method of addressing this problem in the detection of outliers in business surveys taken regularly, and have become pioneers in the study of countermeasures against outliers in business surveys [Hidiroglou-Berthelot method (See Appendix)]. Their method is considered to be a revised version of the quartile method. Specifically, by converting variables for r_{it} or by independently setting the scale of scattering, it aims to (1) detect outliers from both ends of the data distribution, (2) address the problem of "size masking effect," and (3) properly detect outliers even when $(q_2 - q_1)$ and $(q_3 - q_2)$, which are the scales of scattering, are extremely small.

While the above-mentioned three methods are non-parametric methodologies, Smirnov-Grubbs test, a parametric methodology, is sometimes used. However, it is difficult to apply this parametric method when specific distribution, such as normal distribution, cannot be assumed.

Numerous other methodologies have been proposed and studied for the detection of outliers in sample surveys, but they all have pros and cons. Therefore, it is necessary to select and set the optimum method based on the characteristics of the data being analyzed.

(B) The Treatment of Outliers

The next problem is how to treat the detected outliers in estimating the values for the population. There are numerous methods of treating outliers, but they can be classified roughly into the following categories, depending on their characteristics. They are (1) "weight modification," under which the weight of the sample unit is modified, (2) "value modification," under which the value reported by the sample unit is modified, and (3) the combination of the two, under which both the weight and the value reported by the sample are modified[1]. In addition to these three, there is (4) "robust prediction," an estimation method designed to alleviate the impact of the outliers, rather than modifying the outliers themselves (Table 2). The following is an outline of these methodologies.

(a) The Weight Modification Method

Under this method, the weight of the reported value is modified in order to reduce the impact of the outlier. The reported value itself is not modified. For example, when a weighted data from a sample is detected as an outlier because it is extremely large, the weight of the reported value is reduced ("weight reduction") to modify the data downward. The weight of other sample units (i.e., those which are not detected as outliers) is increased to the extent to which the weight of the outlier is reduced.

Here is a simple example. Whereas N_h denotes the number of units in the population in stratum h, in making stratified sampling of n_h, w_h $(=N_h/n_h)$ (the weight of the stratum h) is equal for all samples. Assuming that x_{hi} denotes the value reported by sample i, X_h (the total estimated value for the population in the stratum h) is expressed by Equation (2).

$$\widehat{X_h} = w_h \sum_{i=1}^{n_h} x_{hi} \qquad (2)$$

Out of these samples, let us consider modifying the weight of the samples which have been detected as outliers to w'_h. Assuming that s_1 denotes the group of outliers, that s_2

[1] These methods are called "winsorization".

denotes the group of samples which are not outliers, that n_{h1} denotes the number of samples in s_1, and that n_{h2} ($=n_h - n_{h1}$) denotes the number of samples in s_2, and that the sum of weight is N_h, the total estimated value for the population of X_h after the treatment of the outliers is expressed by Equation (3).

$$\widehat{X_h} = w_h' \sum_{s_1} x_{hi} + \frac{N_h - n_{h1} w_h'}{n_{h2}} \sum_{s_2} x_{hi} \tag{3}$$

This can be regarded as "poststratification" after outliers have been detected. Setting w'_h at 0 (zero) equals expunging the outliers from the computation. The problems of the weight modification method are that (1) it is difficult to set appropriate weight and (2) different weights are given to the samples in the same stratum.

(b) The Value Modification Method

Under this method, rather than modifying the weight of a sample, the reported value is modified in treating an outlier. Major modification methods are (a) to replace the value with that of another sample, (b) when a reported value exceeds the cutoff points set in advance (when the value exceeds the upper limit or falls short of the lower limit), replace it with the value of the cutoff point, (c) assuming that the outlier is not an appropriate data for using in estimation, regard it as a missing value and use missing value imputation, etc. [Chambers and Ren (2004)]. For the last method mentioned here, numerous methods have been devised, including (1) replacing it with a value reported from another sample, (2) using the value reported by the same sample in the past and (3) applying the mean for the stratum (Table 3) [Utsunomiya and Sonoda (2001)].

Since the value modification method does not affect other samples, it is possible to adopt it with relative ease. However, the questions of what value to replace the outlier with and how to set the cutoff points are difficult problems.

(c) The Modification of Both the Weight and Value

In addition to the above two methods, there is a method which is a combination of these two. That is, when the weighted data of a sample is outside of cutoff points (the upper limit=K_U, the lower limit=K_L), instead of replacing the outlier with the value at the cutoff point, give weight of "1" (one) to the portion outside of the cutoff point and add the value to the cutoff point, and use this value [Equation (4)].

$$w_i x_{hi} = \begin{cases} w_i x_{hi} & \text{if } K_L < w_i x_{hi} < K_U \\ K_U + \left(x_{hi} - \frac{K_U}{w_i}\right) & \text{if } K_U \le w_i x_{hi} \\ K_L - \left(\frac{K_L}{w_i} - x_{hi}\right) & \text{if } w_i x_{hi} \le K_L \end{cases} \tag{4}$$

(d) The Method to Alleviate the Impact of Outliers

Under this method, rather than modifying the value reported or the weight of a sample, a statistical model which is robust against the impact of an outlier is used in computing the estimated value to alleviate the impact of an outlier. A simple example is an estimation method using Equation (5). Here, whereas x_i denotes the value reported by sample i, that n denotes the number of samples, that N denotes the number of units in the population and that $\tilde{x}$ denotes a scale robust against the outlier (e.g., the median or the trim mean), X, the total value which is to be estimated, is sought by the following equation.

$$\hat{X} = \sum_1^n x_i + (N - n)\tilde{x} \tag{5}$$

While using the reported values of n number of samples, by using the scale of position $\tilde{x}$ as the estimated value of (N-n) number of non-sample units, this model alleviates the impact of the outlier. Although numerous other statistical models are proposed to be used, they are often said to be too complex in their structure[2]. Consequently, they are seldom used in business surveys.

(5) Actual Cases of Countermeasures Against Outliers

There are not many cases in which outliers are detected and treated based on clear standards in economic statistical surveys taken in Japan. Meanwhile, there are some cases abroad which have introduced countermeasures against outliers (Table 4).

[2] Barnett [1994], Barnett and Lewis[1994]

Their examples include (1) *Survey of Employment, Payrolls and Hours* (SEPH), (Statistics Canada) and (2) *National Construction Industry Wage Rate Survey*, (Statistics Canada) in Canada, (3) *State and Metro Area Employment, Hours & Earnings*, (Bureau of Labor Statistics) in the U.S. and (4) *Consumer Price Index*, (Australian Bureau of Statistics) in Australia.

Actual methods of outlier treatment in these surveys are as follows. *Survey of Employment, Payrolls and Hours* uses Hidiroglou-Berthelot method, quartile method, etc., for outlier detection and the weight modification, etc., for modification. *National Construction Industry Wage Rate Survey* uses "range edit," which adopts "the mean" for the scale of position and "the standard deviation" for the scale of scattering, to detect outliers, which are treated case by case. *State and Metro Area Employment, Hours & Earnings* reduces the impact of outliers through "weight reduction," while *Consumer Price Index* modifies the value of the outlier to the value next in size to the outlier through winsorization. In this manner, the countermeasures against outliers vary from case to case.

3. Incidence of Outliers in *Tankan* and Countermeasures Against Them

(1) The Basic Philosophy on Countermeasures Against Outliers in *Tankan*

Tankan is a sample survey using stratified sampling. Its population comprises private enterprises with paid-in capital of 20 million yen or more (approximately 210,000 enterprises) based on *Establishment and Enterprise Census* of the Ministry of Internal Affairs and Communications. The *Tankan* survey covers approximately 11,000 enterprises as its samples, which are stratified by three variables: (1) the type of business, (2) the amount of paid-in capital and (3) the number of employees.

Tankan basically eliminates the possibility of erroneous responses through layers of micro-editing and macro-editing to verify the accuracy of reported values. On the other hand, although sample stratification is appropriate when it is made, since the population and sample enterprises are fixed for a period of time, rather than drawing samples for each survey, there is a time lag between the time of stratification and when surveys are taken. Consequently, the incidence of stratum jumpers due to rapid growth or the changes in the business formats of enterprises is unavoidable. This can invite the incidence of outliers.

Therefore, in *Tankan*, countermeasures against outliers need to be taken while being mindful of not damaging the following advantages of the statistics.

(A) The Maintenance of Accuracy

Leveling the estimated values for the population aggressively, their rates of change from the previous year or rates of revision can result in a failure to detect an economic turning point or its signs, rendering the statistics unable to meet user-needs at all times. Basically, the responses to survey items obtained from sample enterprises are thought to properly reflect the picture of the population under survey. Therefore, they should be used as they are in compilation, and efforts must be made to detect only the survey items, the impact of which will be extremely large.

(B) The Maintenance of "Easy to Understand" Quality

The "easy to understand" quality of its content, such as the method of calculating the diffusion indices and the definitions of survey items, is an outstanding characteristic of *Tankan*. At the same time, the estimation methods for the values for the populations in quantitative data are also simple and clear, and are easy to understand. Therefore, it is hoped that the countermeasures against outliers are not excessively complex. At the same time, in order to assure transparency, efforts must be made to eliminate the room for arbitrary judgment.

(C) Curbing Compilation Burden

Tankan is a quarterly survey taken in March, June, September and December of each year. In order to capture the latest business conditions swiftly, the findings of the survey are required to be released promptly the day after the final day of the survey period. If the countermeasures against outliers require too much work, the work burden for the compilation of statistics will increase, and can hamper timely execution and release of *Tankan*. Therefore, methods which will not greatly increase the work burden for the statistics compilation are desired.

(2) An Outline of Sample Designs for *Tankan* and Statistics Compilation in Six Enterprise Categories

As was mentioned at the beginning of this paper, for numerical items, such as sales and fixed investment, for which estimates for the values for the population are made, the mean for the stratum is estimated based on the samples, and this mean is multiplied by the number of enterprises in the population to obtain the estimate for the stratum. By adding the sum for each stratum, the overall total is estimated. This value equals the sum of weighted values of values reported by enterprises.

In addition to the total for the entire population, values are computed by enterprise size [based on capitalization, enterprises are classified into large enterprises (the capitalization of 1 billion yen or more), medium-sized enterprises (100 million yen or more but less than 1 billion yen) and small enterprises (20 million yen or more but less than 100 million yen)] and by sector. Of these, of the greatest interest to statistics users are the rates of change from the previous year [Equation (6)] and the rates of revision [Equation (7)] of the estimated values for the population in the six enterprise categories. The six categories are the large enterprises, medium-sized enterprises and small enterprises, which are further classified into the manufacturing and non-manufacturing sectors (hereinafter referred to as "Six Categories base").

Change from the previous year (%)
= [(The estimated value for the population for this fiscal year – the estimated value for the population for the previous fiscal year)/ the estimated value for the population for the previous fiscal year] x 100
(6)

Rate of revision (%)
= [(The estimated value for the population in the present survey – the estimated value for the population in the previous survey)/the estimated value for the population in the previous survey] x 100
(7)

(3) The Concept of Outliers in *Tankan*

As was mentioned earlier, in *Tankan* a great deal of attention is paid to the rate of change from the previous year or the rate of revision from the previous survey, especially on the six-category basis. Therefore, in working with outliers, this paper takes note of the degree of influence of the value reported by an enterprise on the rate of change from the previous year or the rate of revision from the previous survey of the estimated value for the population in the six categories.

Working with outliers within each of the six categories is judged to be appropriate from the standpoint of dealing with the possibility of the difference in enterprise size causing difference in the degree of influence, such as the "size masking effect." At the same time, since there are more than 1,000 sample enterprises in each category, there is no problem in terms of sample size.

Specifically, assuming that t and t – 1 denote the time (in the case of the change from the previous year: this fiscal year and the previous fiscal year, and in the case of the rate of revision: the present survey and the previous survey), that x_{it} denotes the value reported by the ith enterprise, that w_{it} denotes its weight, and that X_{jt} denotes the estimated value for the population in one of the six categories, the degree of influence of the *i*-th enterprise on the rate of change from the previous year or the rate of revision from the previous survey of Category j (j = 1, …, 6) is defined as follows.

Y_{it} = [The margin of change in the weighted data/|the estimated value for the population on the six-category base in the previous fiscal year (previous survey)|] x 100

$$= \frac{w_{it}x_{it} - w_{it-1}x_{it-1}}{|X_{jt-1}|} \times 100 \tag{8}$$

Here, w_{it} and w_{it-1}, the weight at two points in time (t and t – 1) would be equal, if the numbers of sample enterprises are the same. Normally, however, since the number of samples varies from survey to survey, the two weights are different. However, "the margin of change in the weighted data," resulting from the change in weight, does not represent real change in the relevant sample enterprise. Therefore, if there is no response in either of the surveys (at two points in time), "cell mean imputation" to supplement the missing value is applied to arrive at the same number of samples in both surveys. The weight calculated in this manner is expressed as $w_{i.}$. That is to say, that this weight is obtained by dividing the number of enterprises in the population by the number of enterprises, the responses of which were present in at least one of the surveys taken at t and t – 1. As a result of this adjustment, Equation (8) can be expressed as follows.

$$y_{it} = \frac{w_i.(x_{it} - x_{it-1})}{|X_{jt-1}|} \times 100 \tag{9}$$

Based on the distribution of y_{it} (i.e., the degree of influence on the six-category base), the data which greatly diverges from the rest is judged to be an outlier.

(4) Outlier Detection Methods Suitable for *Tankan*

This section will examine outlier detection methods suitable for *Tankan*. A desirable method for the outlier detection for *Tankan* is one which effectively detects only the y_{it} (the degree of influence as described in the previous section) which greatly diverges from other data. Therefore, studies were conducted applying the actual *Tankan* data, in order to see if any of the detection methods discussed in Chapter 2. (4) (A) is suitable for detecting outliers in *Tankan*.

(A) The Application of Existing Outlier Detection Methods

(a) Quartile Method

The distribution of y_{it} (the degree of influence) for the major survey items in *Tankan* (i.e., sales, current profits, net profits for the term, fixed investment amount and software investment amount) for ten surveys (from the March 2007 survey through June 2009 survey) shows that the percentage of enterprises for which y_{it} = 0 (the value reported by the enterprise is identical to the value reported in the previous fiscal year or previous survey) is extremely high, and that the distribution of y_{it} is not necessarily symmetrical (theoretically, when a reported value does not take a negative number, such as sales or fixed investment, the negative side should be a maximum of 100, while the maximum on the positive side is infinite, and therefore the distribution cannot be regarded to be normal (Table 5-1 - 5-10, Figure 1-1 - 1-2). Among small enterprises, in particular, since many of them do not make fixed investment or software investment, there are cases in which the margin of their change is “0” (zero). For example, the amount of software investment in the small enterprises in the manufacturing sector shows that on average y_{it} (the degree of influence) of nearly 90 percent of the firms is "0" (zero) [Figure 1-2 (4)].

In this manner, as the percentage of "0" (zero) for y_{it} (the degree of influence) is very high and the quartile range is often "0" (zero), it is difficult to use as it is the quartile method, which sets the median as the scale of position and the distance between the first quartile and the median, or the third quartile and the median, as the scale of scattering.

(b) The Ratio Method

What follows is an examination of a method using the ratio of the value reported in the present survey to the value reported in the previous survey. Specifically, using the value reported in the preceding survey (which was taken in March 2009) and that reported in the present survey (which was taken in June 2009), the ratio r_{it} (the value reported in the present survey/the value reported in the preceding survey) was obtained for the rate of revision in fiscal 2009 of sales. When r_{it} is outside of the boundary values set in advance (the upper limit: T_U, the lower limit T_L), it is considered to be an outlier. Points observed were as follows (Figure 2-1 - 2-2, Table 6).

- Even when r_{it} (the ratio) diverges greatly from the other data, the degree of its impact y_{it} does not necessarily diverge greatly (Figure 2-1 - 2-2). For example, there are some data the r_{it} (the ratio) of which diverges greatly from the other data but y_{it} (the degree of influence) of which is close to "0" (zero). This means that there is a risk that the data which should not be detected as outliers are being detected as outliers.
- On the other hand, there are data the r_{it} of which is not relatively large, but y_{it} (the degree of influence) of which is relatively large (Figure 2-1 - 2-2). This means that there is a risk that the data which should be detected as outliers are not being detected as such.
- The percentage of data the r_{it} of which exceeds the upper limit tends to rise as the size of enterprise gets smaller. This means that when enterprises are not classified by size, there is a possibility that the size masking effect is being felt [Table 6 (1)].

These observations show that the method using the ratio fails to effectively detect the data y_{it} (the degree of influence) of which diverges greatly. At the same time, since survey items of *Tankan* include items for which the value reported is likely to be "0" (zero) (such as fixed investment and software investment), as well as items for which the value reported can be a negative number (such as current profits and net profits for the term), which makes it difficult to calculate the ratio, the ratio method is not considered to be suitable for *Tankan*.

(c) Hidiroglou-Berthelot Method

In a manner similar to the above, Hidiroglou-Berthelot method (which is described in Appendix) was tested. As in the case of the ratio method, it tried to detect outliers in fiscal 2009 rate of revision of sales obtained in the survey taken in June 2009. The points observed were as follows (Figure 3-1 - 3-2, Table 7).

- As in the case of r_{it}, even when E_{it} (in which variables were converted for r_{it}) greatly diverges from other data, its y_{it} (the degree of influence) does not always diverge greatly (Figure 3-1 - 3-2). For example, there are number of data, the E_{it} of which greatly diverges, but the y_{it} (the degree of influence) of which is close to "0" (zero). This means that there is a risk that data which should not be detected as outliers are being detected as such.
- On the other hand, there are data, E_{it} of which is not relatively large, but the y_{it} of which is relatively large (Figure 3-1 - 3-2). This means that there is a risk that data which should be detected as outliers are not being detected as such.
- With respect to the distribution of E_{it}, there are categories in which both the median and the third quartile are "0" (zero) (Medium-sized enterprises in the manufacturing sector, medium-sized enterprises in the non-manufacturing sector and small enterprises in the manufacturing sector) [Table 7(1)]. As a result, D_U (the scale of scattering) becomes "0" (zero), which makes it impossible to set an appropriate tolerable range. As the upper limit is "0" (zero) in this case, even the data which have changed only slightly are detected as outliers. In fact, nearly 20 percent of the data were detected as outliers [Table 7(3)].

In this manner, even with Hidiroglou-Berthelot method, it is difficult to effectively detect data the y_{it} of which diverges greatly. At the same time, as this method requires that the ratio of the value reported in the present survey to that reported in the previous survey be obtained at first, it is judged to be difficult to apply it for items, such as fixed investment and software investment, for which a high percentage of the values reported is “0” (zero), or for current profits and net profits for the term, which are sometimes in negative numbers.

(d) Hypothesis Test Method

Finally, with respect to the method using hypothesis test, as it is difficult to assume normal distribution or other specific distribution for the values reported in *Tankan* surveys or the degree of influence y_{it}, which is obtained from them, it is not thought appropriate to use this method, which is parametric.

Given these results, the desirable method of effectively detecting outliers in *Tankan* data is to directly apply “range edit” to y_{it} (the degree of influence).

(B) How to View Outlier Detection Suitable for *Tankan*

Based on the studies in the previous section, z_{it} (the indicator of the degree of divergence) is defined, in order to quantitatively grasp the degree of divergence of y_{it} (the degree of influence). Here, in order to address the problem of masking effect, scales robust against outliers are adopted as the scale of position and the scale of scattering. Specifically, with respect to the scale of position, given the fact that the distribution of y_{it} (the degree of influence on the estimated value for the population) is asymmetrical, 1 percentile (d_j^1) and 99 percentile (d_j^{99}) are used. The distance between the 1 percentile and 99 percentile is defined as D_j (the unit of distance between samples) and this value, rather than the quartile range, is used as the scale of scattering.

$$D_j = d_j^{99} - d_j^1 \tag{10}$$

This D_j (the unit of distance between samples) is stable at $D_j > 0$, and therefore can be used as the scale of scattering (Table 8-1 - 8-2). If it should turn out that $D_j = 0$, the mean value for the latest four surveys is to be used. Therefore, Z_{it} is defined as follows.

$$z_{it} = \begin{cases} \frac{y_{it} - d_j^{99}}{D_j} & \text{if } d_j^{99} \leq y_{it} \\ \frac{d_j^1 - y_{it}}{D_j} & \text{if } y_{it} \leq d_j^1 \\ 0 & \text{if } d_j^1 < y_{it} < d_j^{99} \end{cases} \tag{11}$$

This means that the values which are the objects of outlier detection are only those y_{it} (the degree of influence) of which is within 1 percent from the top or within 1 percent from the bottom.

(C) The Method of Determining the Degree of Divergence

Here, the method of judging outliers using z_{it} (the indicator of the degree of divergence) will be discussed. In practice, the standard C is set in advance based on empirical analyses using past data, and when z_{it} exceeds this standard C, it is detected as an outlier (Figure 4).

For the clarity of the detection method, C is to be the same value for all enterprise classifications and survey items. As the indicator of divergence, z_{it} [i.e., D_j, d_j^1, d_j^{99}], is calculated for each of the six categories and for each survey item, it is possible to say that it already reflects the data attributes of each classification and survey item.

(5) The Method of Outlier Treatment

In this paper, in detecting outliers in *Tankan*, in view of *Tankan*'s objectives, priority is given to not to produce too many outliers. That is, as only the data with very large impact are detected as outliers for each survey item, the probability of similar data existing in the population is thought to be low. Moreover, since these outliers are stratum jumpers, they are judged to be put in a wrong stratum as far as a particular survey item is concerned. In other words, the data in the survey item detected as outliers are judged to be non-representative. On the other hand, if the company is not detected as an outlier for other survey items, it is judged to be representative. Because of these considerations, in the case of *Tankan*, even if a value reported by an enterprise is correct, if it is detected as an outlier, it is not used in estimating the value for the population. It is thought that it is appropriate to regard as a missing value the survey item which has been detected as an outlier.

Since 2004, *Tankan* has been using the value reported in the preceding year or the preceding survey ($x_{it} = x_{it-1}$) to make up for a value missing in a survey. Therefore, applying missing value imputation to treat a detected outlier is consistent with this rule. Under this method, y_{it} (the degree of influence) of an outlier on the rate of change from the previous year or the rate of revision from the previous survey of the estimated value for the population (to which the outlier belongs) is "0" (zero). In this manner, the impact of the outlier is completely removed. (Hereafter, this method will be referred to as "Cold Deck Imputation").

4. Empirical Analyses Using Actual Data

In this chapter, the above mentioned methods of detection and treatment of outliers were applied to actual *Tankan* data to detect and treat outliers. With respect to detection, the relationship between the values of the standard C set in advance and the number of data detected as outliers was studied. As for treatment, the data detected as outliers were treated with the methods described in Chapter 3(5), to make trial calculations of the impact on the estimated value for the population.

(1) Data Set

The data set used are from the following five survey items, which are principal items of *Tankan*: sales, current profits, net profits for the term, fixed investment and software investment. The data period covers 10 surveys from the one taken in March 2007 to the one taken in June 2009. Accordingly, there are 90 series (15 series x 6 enterprise categories) for sales, current profits and net profits for the term, and 84 series (14 series x 6 enterprise categories) for the amounts of fixed investment and software investment.

Data set	
Item (5 items)	Sales, current profits, net profits for the term, fixed investment, software investment
Data period (10 surveys)	From March 2007 survey to June 2009 survey Sales, current profits, net profits for the term: 90 series (15 series x 6 categories) Fixed investment, software investment: 84 series (14 series x 6 categories)

(2) Empirical Analyses for Outlier Detection

(The number of data detected)

The detection of outliers was conducted for the above data set. The relationship between the value of standard C and the number of data detected is as follows (Figure 5-1 - 5-10, Table 9).

- When C = 25, a total of 14 outliers were detected for all survey items. This means that on average 1.4 outliers were detected in each survey (Table 9).
- When C = 50, the number of outliers detected declined sharply. One outlier each was detected for fixed investment and software investment (the total of two). Compared with the case of C = 25, only data which have very large impact on the year-on-year change and the rate of revision from the preceding survey were detected as outliers (Table 9).
- When C = 100, only one data for fixed investment was detected as an outlier. At C = 125, no outlier was detected (Table 9).

(Characteristics by Survey Item)

Next, the number of data detected as outliers was examined more closely. With respect to sales, no outlier was detected when C = 25. While y_{it} (the degree of influence on the estimated value for the population) was within a certain range in a stable manner, this is thought to be due to the fact that z_{it} (the indicator of divergence) did not become a large value as D_j (the unit of distance between samples) did not vary greatly in any survey.

For items relating to profits (current profits and net profits for the term), two outliers were detected when C = 25. Unlike sales or fixed investment, responses of enterprises on these items can take negative numbers, which render the absolute value of the estimate for the population relatively small. As a result, y_{it} (the degree of influence) tends to fluctuate. For example, when we examine net profits for the term of the survey taken in March 2009 (large enterprises in the manufacturing sector, the fiscal 2009 figure as compared with the previous year) from Table 5-5(1), partly due to the fact that the actual amount for net profits in the current term (fiscal 2008) was extremely small as compared with normal years, the maximum y_{it} (the degree of influence) is extremely large at “2,291” (which means that the contribution ratio of the relevant data on the rate of change from the preceding year of the net profits for the current term of large enterprises in the manufacturing sector is 2,291 percentage points). However, as many of the other enterprises also show relatively large fluctuations in net profits for the current term, D_j (the unit of distance between samples) is also larger than in other survey items (Table 8-1). As a result, z_{it} (the indicator of the degree of divergence) of the relevant data is within a certain range, suggesting that this data needs not be detected as an outlier. In this manner, even when y_{it} (the degree of influence) fluctuates greatly, as z_{it} (the indicator of divergence) is obtained by dividing it by D_j (the unit of distance between samples), if the relevant number does not diverge from the data distribution, it is not detected as an outlier (thanks to the way the computation mechanism has been devised).

For items on investment (the amounts of fixed investment and software investment), 12 outliers were detected at C = 25, and two at C = 50. Some enterprises among small enterprises in the manufacturing sector show data, the z_{it} (the indicator of divergence) of which is extremely large with respect to the amount of fixed investment, as do some enterprises in the medium-sized enterprises in the non-manufacturing industries with respect to the amount of software investment [Figure 5-7(3), 5-10(2)]. Compared with survey items on profits, the y_{it} of investment items is generally more stable. However, there are cases in which the absolute value of y_{it} (the degree of influence) is near “10,” though these are very rare, and there are also sporadic cases in which it is “between 5 and 10.” Since D_j (the unit of distance between samples) is relatively stable, z_{it} (the degree of divergence) tends to move along with y_{it} (the degree of influence).

These findings of the empirical analyses suggest that while sales and profit items are not very likely to produce outliers, fixed investment and software investment, especially of medium-sized and small enterprises, are more likely to produce outliers. As the number of detected outliers varies depending on the value of C, where the boundary value is set is of great importance in establishing rules for countermeasures against outliers.

(3) Empirical Analyses on the Treatment of Outliers

Next, the method of treating outliers described in Chapter 3(5), “(A) Cold Deck Imputation” was applied to the data detected as outliers to calculate their impact on the estimated value for the population. At the same time, in order to compare impacts of different imputation methods for missing values on the estimated value for the population, figures were obtained for other imputation methods. The first is the "cell mean imputation," in which the mean for the stratum to which the data belong is used (i.e., equal to giving "0" (zero) weight to the value reported and removing it from the computation) (hereafter this method will be referred to as "(B) cell mean imputation). Taking note of the fact that *Tankan* users take note of the rate of change from the previous year or the rate of revision from the previous survey, the second method uses the rate of change from the previous year or the rate of revision from the previous survey of the stratum in which the outliers belong (hereafter this method will be referred to as "(C) the rate of growth imputation). Specifically, under this method, the value reported in the previous year or in the previous survey of the data which is found to be an outlier is multiplied by the rate of change from the previous year or the rate of revision from the previous survey of the stratum as a whole, and the figure is used as the response for this fiscal year or the present survey. Consequently, as the outlier does not influence the rate of change from the previous year or the rate of revision from the previous survey of the relevant stratum, these rates after the treatment of the outlier equal those which are obtained from the responses other than the outlier.

In the following empirical analyses, the cases with the three largest z_{it} (the indicator of the degree of divergence) are assumed to be outliers [the cases (a) - (c) below], and calculations were made to see the extent to which different treatment methods affect the rates of revisions of the six enterprise categories.

(a) June 2007 Survey: Small enterprises in the manufacturing sector, the amount of fixed investment in fiscal 2006 [z_{it} (the indicator of the degree of divergence) = +104.81]

(b) June 2007 Survey: Medium-sized enterprises in the non-manufacturing sector, the amount of software investment in fiscal 2007 [z_{it} (the indicator of the degree of divergence) = +82.18]

(c) September 2008 Survey: Small enterprises in the manufacturing sector, the amount of fixed investment in fiscal 2008 [z_{it} (the indicator of the degree of divergence) = +46.11]

The results of the trial calculations are as follows (Table 10).

(A) In the case of "Cold Deck Imputation," as the impact of the outlier on the rate of revision of the value in the enterprise category was completely eliminated [y_{it} (the degree of influence) is set at "0" (zero)], the rates of the revisions in the six enterprise categories changed by approximately 10 percentage points compared with the estimated values when the outlier was not treated.

(B) In the case of "cell mean imputation," for (a) and (c) in Table 10(2), the signs of y_{it} (the degree of influence) were reversed after the outlier treatment. This means that while the actual values reported by samples were upward revisions from the previous survey, as the mean for the stratum was used, the values were revised downward. In the case of (b), the outlier in question diverged even more greatly from the other data after the outlier treatment.

(C) In the case of "growth rate imputation," in the cases of (a) and (c) in Table 10(2), y_{it} (the degree of divergence) became relatively small [(a) -0.07, (c) +0.04] and compared with the "Cold Deck Imputation," large differences were not observed. However, in the case of (b), y_{it} (the degree of influence) again became relatively large (y_{it} = +7.38) after the outlier treatment. This means one of the data which has been subjected to outlier treatment can again become an outlier.

A comparison of the characteristics of these treatment methods is as follows. The "(A) Cold Deck Imputation" attains the given purpose of eliminating the impact of an outlier. Also, it is simple in practical application. However, regardless of the sign (+ or -), the rate of change from the previous year or the rate of revision from the previous survey (absolute value) calculated by using data other than the outlier is always smaller than the figure obtained by using the outlier. In the "(B) cell mean imputation," when the rate of change from the previous year or the rate of revision from the previous survey of the data, which was detected as an outlier, diverged greatly from that of other data in the same stratum, the use of the cell mean results in a great change in the value reported in the present fiscal year or the present survey from the value reported in the previous year's or previous survey by the relevant enterprise. The "(C) growth rate imputation" does not have an impact on the "rate of change from the previous year or the rate of revision from the previous survey of the stratum as a whole" calculated after removing the outlier. However, when the outlier is greater than values reported by other enterprises in the same stratum, the use of the growth rate imputation can greatly change the estimated value for the population. Moreover, the treatment of survey items, such as current profits, which can be in negative numbers, must be studied further, as it is not always possible to calculate the rate of change of these survey items.

However, since the real picture of the population when the outlier occurred is unknown, it is difficult to choose the optimum method through empirical analyses. In methods using simulations, since the "optimum method" arrived at as a result of the simulations depends on how data for the "assumed" population are set, it is difficult to choose a treatment method which is statistically optimum. However, since the data which have extremely large impact are detected as outliers according to set rules, this paper suggests that it is relatively advantageous to treat outliers than do nothing about them.

5. Conclusions

This paper examined the methods of detecting and treating outliers in a manner suitable for the characteristics of *Tankan* through empirical analyses of actual *Tankan* data. In doing so, while taking into account the existing methods, methods which are suitable for *Tankan* were examined, keeping in mind 1) the accuracy of *Tankan*, 2) the maintenance of its "easy to understand" quality, and 3) the reduction of its work load. Specifically, while the detection method is based on the existing "range edit" method, its scale of position and scale of scattering were defined by using 1 percentile and 99 percentile, so that it can be applied to non-parametric data, for which the incidence of "0" (zero) is very high. This method was chosen so that outliers can be detected in a stable manner. The outliers thus detected are regarded as missing values and should be augmented by applying the missing value imputation, which is already being used for *Tankan*.

The results of the empirical analyses show that the outlier "detection" method shown in this paper is effective in detecting data which have large impact. The "treatment" method is also found to be effective in removing the impact of outliers. These countermeasures against outliers involve relatively limited work burden, which makes it rather easy to introduce them to *Tankan*.

In introducing countermeasures against outliers in the future, however, the following practical problems also need to be addressed.

First, in detecting outliers, one value must be assigned as the standard C. However, it is difficult to set objective standards for judgment in setting the standard C. Hence, taking into account the characteristics of *Tankan* and the objective of taking countermeasures against outliers, and also taking the position of "avoiding the risk of detecting as outliers data which basically are not outliers rather than the risk of failure to detect outliers" (i.e., the avoidance of over-detection rather than under-detection), it appears appropriate to set this standard C in such a way as to limit the number of outliers detected.

Second, the consistency among survey items must be maintained. As outliers are detected for each survey item, only the response which was detected as an outlier needs to be addressed. However, caution is necessary when there are subcategories in the survey item. For example, when the amount of fixed investment and the amount of investment in land, which is a subcategory of the former, are subject to outlier detection, and when an outlier was detected and treated only in one of them, the amount in land investment could become larger than the amount of fixed investment or otherwise produce inconsistency among survey items. Therefore, certain rules are needed. This may include not subjecting the amount of land investment to outlier detection, unless the reported amount of fixed investment is found to be an outlier, in which case both the amounts of fixed investment and land investment are treated in the same manner (e.g., to use the value reported in the previous year or in the previous survey).

Third, the treatment of outliers in future surveys needs to be studied. Since the data detected as an outlier is treated as having been placed in a wrong stratum, it is not desirable to use the relevant sample enterprise for this stratum before the next stratification. That is to say, it is important to eliminate the possibility of unintended impact of repeated outlier treatments on the estimated value for the population.

It is also important to increase the transparency of the statistics as a whole by applying on a continuous basis the countermeasures against outliers which have been introduced, and also disclosing the substance of such countermeasures to outsiders.

Appendix: Hidiroglou-Berthelot method

Revision rate between the current survey and the previous survey, r_{it}, is:

$$r_{it} = \frac{x_{it}}{x_{it-1}}$$

Where x_{it} is data of sample i at the period t. As the distribution of r_{it} is not symmetric, it is pointed out that it is difficult to detect outliers from the left tail of distribution. So, in order to detect outliers from the both tail of distribution r_{it} is converted to S_{it} as follows. Where r_{it}^{M} is median of r_{it}.

$$s_{it} = \begin{cases} 1 - \dfrac{r_{it}^{M}}{r_{it}} & \text{if } 0 < r_{it} < r_{it}^{M} \\ \dfrac{r_{it}}{r_{it}^{M}} - 1 & \text{otherwise } r_{it} \geq r_{it}^{M} \end{cases}$$

Then, considering size effect of sample data E_{it} is defined as follows. Where U is constant in order to adjust size effect:$0 \leq U \leq 1$.

$$E_{it} = s_{it}\{\max(x_{it-1}, x_{it})\}^{U}$$

First, scale of scattering $[D_L,\ D_U]$ is defined as follows. Where E_{it}^{25} is the first quartile; E_{it}^{75} is the third quartile; E_{it}^{M} is median. And, AE_{it}^{M} is defined in order to properly detect outlier when E_{it} concentrate at a certain point and $[E_{it}^{M} - E_{it}^{25}]$ or $[E_{it}^{75} - E_{it}^{M}]$ are extremely small. In practice, A=0.05.

$$D_L = \max\left(E_{it}^{M} - E_{it}^{25}, |AE_{it}^{M}|\right)$$

$$D_U = \max\left(E_{it}^{75} - E_{it}^{M}, |AE_{it}^{M}|\right)$$

Second, set a tolerable range by using constant C. E_{it} is judged to be an outlier when it is outside of this range.

$$\left(\mathrm{E}_{\mathrm{it}}^{\mathrm{M}} - \mathrm{cD}_{\mathrm{L}}, \mathrm{E}_{\mathrm{it}}^{\mathrm{M}} + \mathrm{cD}_{\mathrm{U}}\right)$$

Reference

Belcher, R., (2003). ”Application of the Hidiroglou-Berthelot Method of Outlier Detection for Periodic Business Surveys”. *Proceedings of the Survey Methods Section*, 25-30.

Banim, J., (2000). ”An Assessment of Macro Editing Methods”. *UN/ECE Work Session on Statistical Data Editing*, Working Paper No.7.

Barnett, V., (1994). ”Outlier in Sample Survey”. *Journal of Applied Statistics*, 21, 373-381.

Barnett, V. and Lewis, T., (1994). *Outliers in Statistical Data*, Wiley.

Bernier, J. and Nobrega, K., (1998). ”Outlier Detection in Asymmetric Samples:A Comparison of an Inter-quartile Range Method and a Variation of a Sigma Gap Method”. *Proceedings of the Survey Methods Section*, 137-141.

Chambers, R.L., (1986). ”Outlier Robust Finite Population Estimation”. *Journal of the American Statistical Association*, 81, 1063-1069.

Chambers, R.L. and Ren, R., (2004). ”Outlier Robust Imputation of Survey Data”. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 3336-3344.

Fuller, W. A., (2009). *Sampling Statistic*s, Wiley.

Hidiroglou, M.A., and Berthelot, J.-M., (1986). ”Statistical Editing and Imputation for Periodic Business Surveys”. *Survey Methodology*, 12, 73-83.

Hunt, J.W., (1999). ”Detecting Outliers in the Monthly Retail Trade Survey Using the Hidiroglou-Berthelot Method”. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 539-543.

Lohr, S.L., (1999). *Sampling:Design and Analysis*, Duxbury.

Pfeffermann, D. and Rao, C.R., (2009). *Handbook of Statistics 29*, North Holland.

Utsunomiya and Sonoda (2001). “Methodology for handing missing value in TANKAN”. Working Paper 01-11, Bank of Japan

Table 1: Outliers detections

		Outliers detections
range edit		When x_i exceeds the upper limit or lower limit set in advance, it is detected to be an outlier. Limit is set by mean, median, variance, quartile, etc.
	quartile method	Assuming that q_2 denotes the median, that q_1 denotes the first quartile, that q_3 denotes the third quartile, the scale of position is set as q_2, the scales of scattering as $(q_2 - q_1)$ and $(q_3 - q_2)$, and the tolerable range as $[q_2 - C_L (q_2 - q_1), q_2 + C_U (q_3 - q_2)]$.
	taking note of the ratio	In a survey taken at period t, r_{it}, which is the ratio of the value reported by sample i in the present survey (period t) to the value reported by i in the previous survey (period t - 1), is used to detect outliers.
	Hidiroglou-Berthelot method	Their method is considered to be a revised version of the quartile method. Specifically, by converting variables for r_{it} or by independently setting the scale of scattering, it aims to (1) detect outliers from both ends of the data distribution, (2) address the problem of “size masking effect,” and (3) properly detect outliers even when $(q_2 - q_1)$ and $(q_3 - q_2)$, which are the scales of scattering, are extremely small.
parametric methodology		Sometimes Smirnov-Grubbs test is used.

Table 2: Treatment of Outliers

		Treatment of Outliers
modification		
	weight modification	The weight of the reported value is modified in order to reduce the impact of the outlier. The reported value itself is not modified. For example, when a weighted data from a sample is detected as an outlier because it is extremely large, the weight of the reported value is reduced ("weight reduction") to modify the data downward.
	value modification	The reported value is modified in treating an outlier. Major modification methods are (a) to replace the value with that of another sample, (b) when a reported value exceeds the cutoff points set in advance (when the value exceeds the upper limit or falls short of the lower limit), replace it with the value of the cutoff point, (c) assuming that the outlier is not an appropriate data for using in estimation, regard it as a missing value and use missing value imputation, etc.
	modification of both the weight and value	When the weighted data of a sample is outside of cutoff points (the upper limit=K_U, the lower limit=K_L), instead of replacing the outlier with the value at the cutoff point, give weight of "1" (one) to the portion outside of the cutoff point and add the value to the cutoff point, and use this value.
robust prediction		A statistical model which is robust against the impact of an outlier is used in computing the estimated value to alleviate the impact of an outlier.

Table 3: Imputation

	Imputation
deductive imputation	Using logical relations among the variable.
cell mean imputation	Assuming the missing data are missing completely at random, the average of the value of reporting units is imputed for missing data.
hot-deck imputation	The value of one of the responding units is imputed for missing data. There is several methods; 1) sequential hot-deck imputation, 2) random hot-deck imputation, 3) nearest-neighbor hot-deck imputation.
regression imputation	Predicting the missing value by using a regression of item.
cold-deck imputation	The value of previous survey, such as historical data, is imputed for missing data.
substitution	Choosing a substitute sample when data is missing.
multiple imputation	Missing value is imputed different times. Typically, the same stochastic model is used.

Table 4: Method of other statistics abroad

Statistics	Organization	country	Method
Survey of Employment, Payrolls and Hours(SEPH)	Statistics Canada	Canada	Detected by Hidiroglou-Berthelot method and quartile method. Treatment method is weight modification etc.
National Construction Industry Wage Rate Survey	Statistics Canada	Canada	Uses “range edit,” which adopts “the mean” for the scale of position and “the standard deviation” for the scale of scattering, to detect outliers, which are treated case by case.
State and Metro Area Employment, Hours, & Earnings	Bureau of Labor Statistics	America	Reduces the impact of outliers through “weight reduction”.
Consumer Price Index	Australian Bureau of Statistics	Australia	Modifies the value of the outlier to the value next in size to the outlier through winsorization.

Table 5-1: Distributions of y_{it} to population estimates

(1) Sales: Large enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1245	0.138	-0.138	0.001	662 (53.2)	365 (29.3)	218 (17.5)
	June survey	revision rate (FY 2006)	1245	0.238	-0.034	0.001	668 (53.7)	177 (14.2)	400 (32.1)
		revision rate (FY 2007)	1245	0.287	-0.049	0.002	609 (48.9)	318 (25.5)	318 (25.5)
	Sept. survey	revision rate (FY 2007)	1239	0.155	-0.027	0.001	312 (25.2)	687 (55.4)	240 (19.4)
	Dec. survey	revision rate (FY 2007)	1235	0.107	-0.048	0.001	511 (41.4)	270 (21.9)	454 (36.8)
FY 2008									
	March survey	revision rate (FY 2007)	1233	0.129	-0.246	0.000	318 (25.8)	597 (48.4)	318 (25.8)
		year-on-year rate (FY 2008)	1226	0.205	-0.110	0.002	609 (49.7)	394 (32.1)	223 (18.2)
	June survey	revision rate (FY 2007)	1226	0.096	-0.054	0.000	587 (47.9)	164 (13.4)	475 (38.7)
		revision rate (FY 2008)	1226	0.220	-0.108	0.002	553 (45.1)	306 (25.0)	367 (29.9)
	Sept. survey	revision rate (FY 2008)	1219	0.320	-0.055	0.001	223 (18.3)	679 (55.7)	317 (26.0)
	Dec. survey	revision rate (FY 2008)	1215	0.092	-0.438	-0.003	261 (21.5)	198 (16.3)	756 (62.2)
FY 2009									
	March survey	revision rate (FY 2008)	1214	0.066	-0.419	-0.007	79 (6.5)	256 (21.1)	879 (72.4)
		year-on-year rate (FY 2009)	1207	0.088	-0.405	-0.005	202 (16.7)	353 (29.2)	652 (54.0)
	June survey	revision rate (FY 2008)	1208	0.099	-0.114	-0.001	368 (30.5)	151 (12.5)	689 (57.0)
		revision rate (FY 2009)	1208	0.120	-0.653	-0.008	274 (22.7)	259 (21.4)	675 (55.9)

(2) Sales: Medium-sized enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1236	0.245	-0.214	0.002	745 (60.3)	98 (7.9)	393 (31.8)
	June survey	revision rate (FY 2006)	1236	0.205	-0.071	0.001	635 (51.4)	246 (19.9)	355 (28.7)
		revision rate (FY 2007)	1236	0.220	-0.080	0.001	395 (32.0)	549 (44.4)	292 (23.6)
	Sept. survey	revision rate (FY 2007)	1228	0.073	-0.233	0.000	463 (37.7)	345 (28.1)	420 (34.2)
	Dec. survey	revision rate (FY 2007)	1219	0.067	-0.030	0.000	573 (47.0)	138 (11.3)	508 (41.7)
FY 2008									
	March survey	revision rate (FY 2007)	1220	0.141	-0.514	0.000	467 (38.3)	283 (23.2)	470 (38.5)
		year-on-year rate (FY 2008)	1201	0.157	-0.097	0.002	741 (61.7)	92 (7.7)	368 (30.6)
	June survey	revision rate (FY 2007)	1203	0.200	-0.066	0.001	601 (50.0)	239 (19.9)	363 (30.2)
		revision rate (FY 2008)	1203	0.199	-0.074	0.001	403 (33.5)	506 (42.1)	294 (24.4)
	Sept. survey	revision rate (FY 2008)	1198	0.133	-0.088	0.000	401 (33.5)	318 (26.5)	479 (40.0)
	Dec. survey	revision rate (FY 2008)	1193	0.029	-0.145	-0.002	381 (31.9)	116 (9.7)	696 (58.3)
FY 2009									
	March survey	revision rate (FY 2008)	1189	0.048	-0.181	-0.005	183 (15.4)	161 (13.5)	845 (71.1)
		year-on-year rate (FY 2009)	1178	0.205	-0.291	-0.007	272 (23.1)	80 (6.8)	826 (70.1)
	June survey	revision rate (FY 2008)	1178	0.133	-0.111	0.000	455 (38.6)	214 (18.2)	509 (43.2)
		revision rate (FY 2009)	1178	0.078	-0.192	-0.002	268 (22.8)	418 (35.5)	492 (41.8)

(3) Sales: Small enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	2035	0.138	-0.116	0.001	1085 (53.3)	182 (8.9)	768 (37.7)
	June survey	revision rate (FY 2006)	2036	0.074	-0.065	0.001	953 (46.8)	592 (29.1)	491 (24.1)
		revision rate (FY 2007)	2036	0.117	-0.054	0.001	672 (33.0)	821 (40.3)	543 (26.7)
	Sept. survey	revision rate (FY 2007)	2017	0.211	-0.064	0.000	752 (37.3)	549 (27.2)	716 (35.5)
	Dec. survey	revision rate (FY 2007)	2005	0.156	-0.185	0.000	857 (42.7)	309 (15.4)	839 (41.8)
FY 2008									
	March survey	revision rate (FY 2007)	2007	0.075	-0.344	0.000	767 (38.2)	490 (24.4)	750 (37.4)
		year-on-year rate (FY 2008)	1981	0.340	-0.099	0.001	997 (50.3)	215 (10.9)	769 (38.8)
	June survey	revision rate (FY 2007)	1981	0.070	-0.101	0.000	881 (44.5)	555 (28.0)	545 (27.5)
		revision rate (FY 2008)	1981	0.157	-0.104	0.000	601 (30.3)	764 (38.6)	616 (31.1)
	Sept. survey	revision rate (FY 2008)	1959	0.126	-0.135	0.000	627 (32.0)	465 (23.7)	867 (44.3)
	Dec. survey	revision rate (FY 2008)	1947	0.063	-0.229	-0.001	635 (32.6)	244 (12.5)	1068 (54.9)
FY 2009									
	March survey	revision rate (FY 2008)	1933	0.084	-0.101	-0.002	416 (21.5)	347 (18.0)	1170 (60.5)
		year-on-year rate (FY 2009)	1911	0.115	-0.412	-0.005	421 (22.0)	170 (8.9)	1320 (69.1)
	June survey	revision rate (FY 2008)	1916	0.125	-0.148	0.000	731 (38.2)	478 (24.9)	707 (36.9)
		revision rate (FY 2009)	1916	0.133	-0.269	-0.002	445 (23.2)	624 (32.6)	847 (44.2)

Table 5-2: Distributions of y_{it} to population estimates

(1) Sales: Large enterprises of nonmanufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007											
March survey	year-on-year rate (FY 2007)	1222	0.281	-0.205	0.001	582	(47.6)	382	(31.3)	258	(21.1)
June survey	revision rate (FY 2006)	1223	0.083	-0.123	0.000	619	(50.6)	225	(18.4)	379	(31.0)
	revision rate (FY 2007)	1223	0.091	-0.252	0.001	537	(43.9)	337	(27.6)	349	(28.5)
Sept. survey	revision rate (FY 2007)	1217	0.055	-0.101	0.000	256	(21.0)	720	(59.2)	241	(19.8)
Dec. survey	revision rate (FY 2007)	1212	0.140	-0.070	0.001	443	(36.6)	313	(25.8)	456	(37.6)
FY 2008											
March survey	revision rate (FY 2007)	1211	0.071	-0.110	0.000	290	(23.9)	530	(43.8)	391	(32.3)
	year-on-year rate (FY 2008)	1199	0.386	-0.083	0.001	543	(45.3)	389	(32.4)	267	(22.3)
June survey	revision rate (FY 2007)	1199	0.145	-0.024	0.001	528	(44.0)	210	(17.5)	461	(38.4)
	revision rate (FY 2008)	1199	0.266	-0.070	0.002	459	(38.3)	357	(29.8)	383	(31.9)
Sept. survey	revision rate (FY 2008)	1189	0.192	-0.102	0.001	183	(15.4)	680	(57.2)	326	(27.4)
Dec. survey	revision rate (FY 2008)	1186	0.241	-0.177	0.000	329	(27.7)	270	(22.8)	587	(49.5)
FY 2009											
March survey	revision rate (FY 2008)	1182	0.218	-0.690	-0.003	169	(14.3)	387	(32.7)	626	(53.0)
	year-on-year rate (FY 2009)	1171	0.103	-0.612	-0.004	308	(26.3)	395	(33.7)	468	(40.0)
June survey	revision rate (FY 2008)	1172	0.137	-0.414	-0.002	356	(30.4)	197	(16.8)	619	(52.8)
	revision rate (FY 2009)	1172	0.287	-0.913	-0.006	305	(26.0)	320	(27.3)	547	(46.7)

(2) Sales: Medium-sized enterprises of nonmanufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007											
March survey	year-on-year rate (FY 2007)	1690	0.378	-0.052	0.002	916	(54.2)	200	(11.8)	574	(34.0)
June survey	revision rate (FY 2006)	1692	0.484	-0.254	0.000	772	(45.6)	399	(23.6)	521	(30.8)
	revision rate (FY 2007)	1692	0.397	-0.242	0.001	539	(31.9)	718	(42.4)	435	(25.7)
Sept. survey	revision rate (FY 2007)	1675	0.117	-0.407	0.000	493	(29.4)	551	(32.9)	631	(37.7)
Dec. survey	revision rate (FY 2007)	1660	0.089	-0.085	0.000	656	(39.5)	278	(16.7)	726	(43.7)
FY 2008											
March survey	revision rate (FY 2007)	1652	0.078	-0.170	0.000	551	(33.4)	437	(26.5)	664	(40.2)
	year-on-year rate (FY 2008)	1636	0.174	-0.359	0.001	874	(53.4)	193	(11.8)	569	(34.8)
June survey	revision rate (FY 2007)	1636	0.302	-0.069	0.000	717	(43.8)	385	(23.5)	534	(32.6)
	revision rate (FY 2008)	1636	0.312	-0.171	0.001	464	(28.4)	686	(41.9)	486	(29.7)
Sept. survey	revision rate (FY 2008)	1622	0.693	-0.248	0.000	426	(26.3)	538	(33.2)	658	(40.6)
Dec. survey	revision rate (FY 2008)	1618	0.161	-0.172	-0.001	502	(31.0)	269	(16.6)	847	(52.3)
FY 2009											
March survey	revision rate (FY 2008)	1610	0.077	-0.228	-0.002	346	(21.5)	346	(21.5)	918	(57.0)
	year-on-year rate (FY 2009)	1585	0.200	-0.682	-0.003	528	(33.3)	218	(13.8)	839	(52.9)
June survey	revision rate (FY 2008)	1586	0.159	-0.347	0.000	568	(35.8)	371	(23.4)	647	(40.8)
	revision rate (FY 2009)	1586	0.157	-0.368	-0.001	383	(24.1)	608	(38.3)	595	(37.5)

(3) Sales: Small enterprises of nonmanufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007											
March survey	year-on-year rate (FY 2007)	3457	0.241	-0.089	0.000	1568	(45.4)	543	(15.7)	1346	(38.9)
June survey	revision rate (FY 2006)	3458	0.066	-0.234	0.000	1401	(40.5)	1128	(32.6)	929	(26.9)
	revision rate (FY 2007)	3458	0.094	-0.234	0.000	984	(28.5)	1555	(45.0)	919	(26.6)
Sept. survey	revision rate (FY 2007)	3418	0.099	-0.064	0.000	1022	(29.9)	1183	(34.6)	1213	(35.5)
Dec. survey	revision rate (FY 2007)	3368	0.104	-0.156	0.000	1297	(38.5)	688	(20.4)	1383	(41.1)
FY 2008											
March survey	revision rate (FY 2007)	3406	0.142	-0.063	0.000	1194	(35.1)	974	(28.6)	1238	(36.3)
	year-on-year rate (FY 2008)	3360	0.082	-0.118	0.000	1499	(44.6)	518	(15.4)	1343	(40.0)
June survey	revision rate (FY 2007)	3363	0.062	-0.113	0.000	1368	(40.7)	1081	(32.1)	914	(27.2)
	revision rate (FY 2008)	3363	0.104	-0.114	0.000	906	(26.9)	1445	(43.0)	1012	(30.1)
Sept. survey	revision rate (FY 2008)	3318	0.205	-0.072	0.000	932	(28.1)	1057	(31.9)	1329	(40.1)
Dec. survey	revision rate (FY 2008)	3272	0.123	-0.224	0.000	1139	(34.8)	655	(20.0)	1478	(45.2)
FY 2009											
March survey	revision rate (FY 2008)	3331	0.125	-0.247	-0.001	866	(26.0)	808	(24.3)	1657	(49.7)
	year-on-year rate (FY 2009)	3284	0.129	-0.327	-0.002	947	(28.8)	461	(14.0)	1876	(57.1)
June survey	revision rate (FY 2008)	3285	0.084	-0.131	0.000	1213	(36.9)	1014	(30.9)	1058	(32.2)
	revision rate (FY 2009)	3285	0.163	-0.121	0.000	832	(25.3)	1243	(37.8)	1210	(36.8)

Table 5-3: Distributions of y_{it} to population estimates

(1) Current profits: Large enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007												
	March survey	year-on-year rate (FY 2007)	1245	0.761	-0.761	0.000	567	(45.5)	376	(30.2)	302	(24.3)
	June survey	revision rate (FY 2006)	1245	0.563	-0.843	0.002	700	(56.2)	171	(13.7)	374	(30.0)
		revision rate (FY 2007)	1245	0.621	-0.842	0.002	531	(42.7)	252	(20.2)	462	(37.1)
	Sept. survey	revision rate (FY 2007)	1239	0.631	-0.263	0.002	316	(25.5)	690	(55.7)	233	(18.8)
	Dec. survey	revision rate (FY 2007)	1235	0.320	-0.381	0.001	475	(38.5)	263	(21.3)	497	(40.2)
FY 2008												
	March survey	revision rate (FY 2007)	1233	0.240	-0.370	-0.001	277	(22.5)	585	(47.4)	371	(30.1)
		year-on-year rate (FY 2008)	1226	0.592	-0.381	0.000	505	(41.2)	399	(32.5)	322	(26.3)
	June survey	revision rate (FY 2007)	1226	0.744	-0.736	-0.002	507	(41.4)	171	(13.9)	548	(44.7)
		revision rate (FY 2008)	1226	0.744	-3.284	-0.009	412	(33.6)	257	(21.0)	557	(45.4)
	Sept. survey	revision rate (FY 2008)	1219	1.030	-0.310	0.000	223	(18.3)	665	(54.6)	331	(27.2)
	Dec. survey	revision rate (FY 2008)	1215	0.706	-2.326	-0.013	323	(26.6)	203	(16.7)	689	(56.7)
FY 2009												
	March survey	revision rate (FY 2008)	1214	0.552	-3.183	-0.042	137	(11.3)	247	(20.3)	830	(68.4)
		year-on-year rate (FY 2009)	1207	2.352	-1.710	-0.016	412	(34.1)	371	(30.7)	424	(35.1)
	June survey	revision rate (FY 2008)	1208	3.129	-1.929	0.002	600	(49.7)	150	(12.4)	458	(37.9)
		revision rate (FY 2009)	1208	3.555	-9.704	-0.019	495	(41.0)	211	(17.5)	502	(41.6)

(2) Current profits: Medium-sized enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007												
	March survey	year-on-year rate (FY 2007)	1234	0.752	-0.744	0.001	653	(52.9)	138	(11.2)	443	(35.9)
	June survey	revision rate (FY 2006)	1235	0.712	-0.485	0.003	595	(48.2)	258	(20.9)	382	(30.9)
		revision rate (FY 2007)	1235	0.265	-0.350	0.000	402	(32.6)	437	(35.4)	396	(32.1)
	Sept. survey	revision rate (FY 2007)	1227	0.705	-0.411	0.001	411	(33.5)	349	(28.4)	467	(38.1)
	Dec. survey	revision rate (FY 2007)	1218	0.397	-0.156	-0.001	515	(42.3)	146	(12.0)	557	(45.7)
FY 2008												
	March survey	revision rate (FY 2007)	1219	0.277	-0.482	-0.002	416	(34.1)	271	(22.2)	532	(43.6)
		year-on-year rate (FY 2008)	1199	0.476	-0.823	0.003	642	(53.5)	130	(10.8)	427	(35.6)
	June survey	revision rate (FY 2007)	1201	0.359	-1.452	-0.002	525	(43.7)	239	(19.9)	437	(36.4)
		revision rate (FY 2008)	1201	0.397	-0.747	-0.002	387	(32.2)	393	(32.7)	421	(35.1)
	Sept. survey	revision rate (FY 2008)	1196	0.434	-0.570	-0.003	362	(30.3)	323	(27.0)	511	(42.7)
	Dec. survey	revision rate (FY 2008)	1191	0.479	-0.810	-0.011	380	(31.9)	119	(10.0)	692	(58.1)
FY 2009												
	March survey	revision rate (FY 2008)	1187	0.498	-1.731	-0.030	227	(19.1)	183	(15.4)	777	(65.5)
		year-on-year rate (FY 2009)	1176	0.982	-2.107	-0.017	514	(43.7)	116	(9.9)	546	(46.4)
	June survey	revision rate (FY 2008)	1176	0.495	-1.449	-0.001	524	(44.6)	214	(18.2)	438	(37.2)
		revision rate (FY 2009)	1176	1.066	-1.181	-0.002	422	(35.9)	342	(29.1)	412	(35.0)

(3) Current profits: Small enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007												
	March survey	year-on-year rate (FY 2007)	2030	0.708	-0.516	0.004	976	(48.1)	393	(19.4)	661	(32.6)
	June survey	revision rate (FY 2006)	2031	0.230	-0.279	0.002	817	(40.2)	610	(30.0)	604	(29.7)
		revision rate (FY 2007)	2031	0.247	-0.421	-0.001	620	(30.5)	699	(34.4)	712	(35.1)
	Sept. survey	revision rate (FY 2007)	2014	0.383	-0.341	-0.002	595	(29.5)	621	(30.8)	798	(39.6)
	Dec. survey	revision rate (FY 2007)	2002	0.260	-1.514	-0.002	760	(38.0)	377	(18.8)	865	(43.2)
FY 2008												
	March survey	revision rate (FY 2007)	2004	0.373	-0.669	-0.003	611	(30.5)	600	(29.9)	793	(39.6)
		year-on-year rate (FY 2008)	1977	7.769	-0.437	0.008	936	(47.3)	394	(19.9)	647	(32.7)
	June survey	revision rate (FY 2007)	1977	0.304	-0.466	-0.001	717	(36.3)	600	(30.3)	660	(33.4)
		revision rate (FY 2008)	1977	0.412	-0.720	-0.004	538	(27.2)	642	(32.5)	797	(40.3)
	Sept. survey	revision rate (FY 2008)	1957	0.580	-2.997	-0.005	525	(26.8)	564	(28.8)	868	(44.4)
	Dec. survey	revision rate (FY 2008)	1945	0.927	-0.828	-0.008	620	(31.9)	335	(17.2)	990	(50.9)
FY 2009												
	March survey	revision rate (FY 2008)	1931	0.436	-1.852	-0.018	406	(21.0)	426	(22.1)	1099	(56.9)
		year-on-year rate (FY 2009)	1909	3.594	-4.152	-0.017	697	(36.5)	311	(16.3)	901	(47.2)
	June survey	revision rate (FY 2008)	1913	1.088	-0.904	-0.001	705	(36.9)	545	(28.5)	663	(34.7)
		revision rate (FY 2009)	1913	2.081	-13.56	-0.019	605	(31.6)	528	(27.6)	780	(40.8)

Table 5-4: Distributions of y_{it} to population estimates

(1) Current profits: Large enterprises of nonmanufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007								
March survey	year-on-year rate (FY 2007)	1221	0.226	-0.807	-0.001	474 (38.8)	412 (33.7)	335 (27.4)
June survey	revision rate (FY 2006)	1222	0.398	-0.180	0.003	687 (56.2)	200 (16.4)	335 (27.4)
	revision rate (FY 2007)	1222	0.451	-0.250	0.003	540 (44.2)	297 (24.3)	385 (31.5)
Sept. survey	revision rate (FY 2007)	1216	0.433	-1.633	0.000	268 (22.0)	711 (58.5)	237 (19.5)
Dec. survey	revision rate (FY 2007)	1211	0.775	-0.662	0.001	475 (39.2)	289 (23.9)	447 (36.9)
FY 2008								
March survey	revision rate (FY 2007)	1210	0.258	-0.683	-0.003	295 (24.4)	519 (42.9)	396 (32.7)
	year-on-year rate (FY 2008)	1198	0.160	-0.539	0.001	474 (39.6)	394 (32.9)	330 (27.5)
June survey	revision rate (FY 2007)	1198	0.990	-0.131	0.003	587 (49.0)	195 (16.3)	416 (34.7)
	revision rate (FY 2008)	1198	0.983	-1.265	-0.001	469 (39.1)	290 (24.2)	439 (36.6)
Sept. survey	revision rate (FY 2008)	1188	0.989	-1.232	-0.004	187 (15.7)	662 (55.7)	339 (28.5)
Dec. survey	revision rate (FY 2008)	1185	0.647	-1.223	-0.008	386 (32.6)	263 (22.2)	536 (45.2)
FY 2009								
March survey	revision rate (FY 2008)	1181	1.991	-0.704	-0.007	239 (20.2)	375 (31.8)	567 (48.0)
	year-on-year rate (FY 2009)	1172	1.239	-2.429	-0.005	384 (32.8)	405 (34.6)	383 (32.7)
June survey	revision rate (FY 2008)	1173	0.649	-1.465	-0.005	584 (49.8)	178 (15.2)	411 (35.0)
	revision rate (FY 2009)	1173	1.979	-2.532	-0.007	416 (35.5)	269 (22.9)	488 (41.6)

(2) Current profits: Medium-sized enterprises of nonmanufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007								
March survey	year-on-year rate (FY 2007)	1689	0.472	-0.582	0.005	878 (52.0)	254 (15.0)	557 (33.0)
June survey	revision rate (FY 2006)	1691	0.660	-0.298	0.003	810 (47.9)	365 (21.6)	516 (30.5)
	revision rate (FY 2007)	1691	0.448	-0.414	0.000	572 (33.8)	568 (33.6)	551 (32.6)
Sept. survey	revision rate (FY 2007)	1674	0.238	-0.290	-0.001	512 (30.6)	562 (33.6)	600 (35.8)
Dec. survey	revision rate (FY 2007)	1659	1.400	-0.523	0.000	673 (40.6)	253 (15.3)	733 (44.2)
FY 2008								
March survey	revision rate (FY 2007)	1651	0.403	-0.592	-0.001	534 (32.3)	430 (26.0)	687 (41.6)
	year-on-year rate (FY 2008)	1635	1.070	-0.676	0.005	835 (51.1)	222 (13.6)	578 (35.4)
June survey	revision rate (FY 2007)	1635	0.175	-0.794	0.001	752 (46.0)	329 (20.1)	554 (33.9)
	revision rate (FY 2008)	1635	0.337	-1.049	-0.002	493 (30.2)	553 (33.8)	589 (36.0)
Sept. survey	revision rate (FY 2008)	1621	0.684	-1.930	-0.005	410 (25.3)	527 (32.5)	684 (42.2)
Dec. survey	revision rate (FY 2008)	1617	0.858	-1.687	-0.005	576 (35.6)	260 (16.1)	781 (48.3)
FY 2009								
March survey	revision rate (FY 2008)	1609	0.377	-0.455	-0.011	425 (26.4)	345 (21.4)	839 (52.1)
	year-on-year rate (FY 2009)	1585	1.081	-0.930	0.000	692 (43.7)	260 (16.4)	633 (39.9)
June survey	revision rate (FY 2008)	1586	1.900	-3.085	-0.001	742 (46.8)	339 (21.4)	505 (31.8)
	revision rate (FY 2009)	1586	3.971	-0.480	0.002	529 (33.4)	516 (32.5)	541 (34.1)

(3) Current profits: Small enterprises of nonmanufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007								
March survey	year-on-year rate (FY 2007)	3448	1.982	-0.687	0.002	1494 (43.3)	935 (27.1)	1019 (29.6)
June survey	revision rate (FY 2006)	3449	0.821	-0.410	0.001	1236 (35.8)	1231 (35.7)	982 (28.5)
	revision rate (FY 2007)	3449	0.769	-1.117	-0.001	964 (28.0)	1423 (41.3)	1062 (30.8)
Sept. survey	revision rate (FY 2007)	3411	1.101	-0.363	0.000	908 (26.6)	1316 (38.6)	1187 (34.8)
Dec. survey	revision rate (FY 2007)	3362	0.368	-0.343	0.000	1147 (34.1)	877 (26.1)	1338 (39.8)
FY 2008								
March survey	revision rate (FY 2007)	3399	0.388	-0.458	-0.001	969 (28.5)	1139 (33.5)	1291 (38.0)
	year-on-year rate (FY 2008)	3351	0.723	-1.114	0.001	1438 (42.9)	912 (27.2)	1001 (29.9)
June survey	revision rate (FY 2007)	3355	0.357	-0.536	0.001	1154 (34.4)	1144 (34.1)	1057 (31.5)
	revision rate (FY 2008)	3355	0.377	-0.801	-0.001	898 (26.8)	1300 (38.7)	1157 (34.5)
Sept. survey	revision rate (FY 2008)	3308	0.830	-0.300	-0.001	801 (24.2)	1237 (37.4)	1270 (38.4)
Dec. survey	revision rate (FY 2008)	3265	0.454	-0.869	-0.003	1009 (30.9)	780 (23.9)	1476 (45.2)
FY 2009								
March survey	revision rate (FY 2008)	3324	0.772	-0.953	-0.005	834 (25.1)	981 (29.5)	1509 (45.4)
	year-on-year rate (FY 2009)	3276	1.082	-0.804	0.002	1257 (38.4)	795 (24.3)	1224 (37.4)
June survey	revision rate (FY 2008)	3277	0.792	-0.430	0.001	1191 (36.3)	1109 (33.8)	977 (29.8)
	revision rate (FY 2009)	3277	0.666	-0.640	-0.001	999 (30.5)	1193 (36.4)	1085 (33.1)

Table 5-5: Distributions of y_{it} to population estimates

(1) Net income: Large enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1245	3.060	-2.278	0.002	528 (42.4)	361 (29.0)	356 (28.6)
	June survey	revision rate (FY 2006)	1245	0.567	-1.471	-0.002	625 (50.2)	168 (13.5)	452 (36.3)
		revision rate (FY 2007)	1245	0.904	-1.434	0.004	554 (44.5)	234 (18.8)	457 (36.7)
	Sept. survey	revision rate (FY 2007)	1239	0.591	-0.694	0.002	311 (25.1)	685 (55.3)	243 (19.6)
	Dec. survey	revision rate (FY 2007)	1235	1.652	-0.244	0.002	476 (38.5)	258 (20.9)	501 (40.6)
FY 2008									
	March survey	revision rate (FY 2007)	1233	0.359	-0.955	-0.003	264 (21.4)	588 (47.7)	381 (30.9)
		year-on-year rate (FY 2008)	1226	0.992	-1.202	0.002	492 (40.1)	387 (31.6)	347 (28.3)
	June survey	revision rate (FY 2007)	1226	1.203	-0.953	-0.005	439 (35.8)	173 (14.1)	614 (50.1)
		revision rate (FY 2008)	1226	1.176	-2.985	-0.010	414 (33.8)	259 (21.1)	553 (45.1)
	Sept. survey	revision rate (FY 2008)	1219	0.947	-0.306	0.000	218 (17.9)	669 (54.9)	332 (27.2)
	Dec. survey	revision rate (FY 2008)	1215	0.653	-2.452	-0.016	300 (24.7)	199 (16.4)	716 (58.9)
FY 2009									
	March survey	revision rate (FY 2008)	1214	0.519	-6.155	-0.082	110 (9.1)	247 (20.3)	857 (70.6)
		year-on-year rate (FY 2009)	1207	2291	-615.2	6.506	460 (38.1)	366 (30.3)	381 (31.6)
	June survey	revision rate (FY 2008)	1208	499.0	-776.0	-11.30	439 (36.3)	147 (12.2)	622 (51.5)
		revision rate (FY 2009)	1208	13.05	-11.62	-0.017	521 (43.1)	196 (16.2)	491 (40.6)

(2) Net income: Medium-sized enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1226	1.620	-1.149	0.004	643 (52.4)	139 (11.3)	444 (36.2)
	June survey	revision rate (FY 2006)	1228	2.302	-0.470	0.002	532 (43.3)	254 (20.7)	442 (36.0)
		revision rate (FY 2007)	1228	0.450	-0.705	0.000	404 (32.9)	423 (34.4)	401 (32.7)
	Sept. survey	revision rate (FY 2007)	1222	1.233	-0.426	0.001	413 (33.8)	349 (28.6)	460 (37.6)
	Dec. survey	revision rate (FY 2007)	1213	0.750	-0.337	-0.001	526 (43.4)	150 (12.4)	537 (44.3)
FY 2008									
	March survey	revision rate (FY 2007)	1215	1.221	-0.732	-0.004	399 (32.8)	289 (23.8)	527 (43.4)
		year-on-year rate (FY 2008)	1195	1.218	-1.030	0.009	643 (53.8)	132 (11.0)	420 (35.1)
	June survey	revision rate (FY 2007)	1198	0.598	-2.271	-0.006	476 (39.7)	236 (19.7)	486 (40.6)
		revision rate (FY 2008)	1198	0.472	-0.862	-0.003	374 (31.2)	381 (31.8)	443 (37.0)
	Sept. survey	revision rate (FY 2008)	1193	0.480	-0.653	-0.005	350 (29.3)	332 (27.8)	511 (42.8)
	Dec. survey	revision rate (FY 2008)	1188	0.454	-0.931	-0.016	375 (31.6)	129 (10.9)	684 (57.6)
FY 2009									
	March survey	revision rate (FY 2008)	1185	0.781	-2.591	-0.053	222 (18.7)	197 (16.6)	766 (64.6)
		year-on-year rate (FY 2009)	1174	11.88	-8.219	0.047	550 (46.8)	123 (10.5)	501 (42.7)
	June survey	revision rate (FY 2008)	1174	5.490	-26.73	-0.102	445 (37.9)	216 (18.4)	513 (43.7)
		revision rate (FY 2009)	1174	3.612	-2.774	-0.002	432 (36.8)	326 (27.8)	416 (35.4)

(3) Net income: Small enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	2026	0.995	-2.595	0.007	973 (48.0)	412 (20.3)	641 (31.6)
	June survey	revision rate (FY 2006)	2027	0.733	-4.533	-0.002	747 (36.9)	658 (32.5)	622 (30.7)
		revision rate (FY 2007)	2027	0.731	-0.387	0.000	616 (30.4)	739 (36.5)	672 (33.2)
	Sept. survey	revision rate (FY 2007)	2010	0.366	-0.612	-0.002	577 (28.7)	667 (33.2)	766 (38.1)
	Dec. survey	revision rate (FY 2007)	1998	1.310	-1.413	-0.003	704 (35.2)	423 (21.2)	871 (43.6)
FY 2008									
	March survey	revision rate (FY 2007)	2000	0.431	-1.006	-0.006	583 (29.2)	611 (30.6)	806 (40.3)
		year-on-year rate (FY 2008)	1972	7.097	-0.921	0.013	940 (47.7)	419 (21.2)	613 (31.1)
	June survey	revision rate (FY 2007)	1972	0.548	-0.670	-0.003	650 (33.0)	633 (32.1)	689 (34.9)
		revision rate (FY 2008)	1972	0.619	-1.172	-0.005	530 (26.9)	674 (34.2)	768 (38.9)
	Sept. survey	revision rate (FY 2008)	1954	0.980	-1.823	-0.006	518 (26.5)	575 (29.4)	861 (44.1)
	Dec. survey	revision rate (FY 2008)	1942	0.446	-0.981	-0.013	601 (30.9)	354 (18.2)	987 (50.8)
FY 2009									
	March survey	revision rate (FY 2008)	1928	0.647	-3.322	-0.033	385 (20.0)	440 (22.8)	1103 (57.2)
		year-on-year rate (FY 2009)	1905	14.87	-14.93	-0.013	748 (39.3)	320 (16.8)	837 (43.9)
	June survey	revision rate (FY 2008)	1909	10.35	-7.823	-0.012	671 (35.1)	564 (29.5)	674 (35.3)
		revision rate (FY 2009)	1909	5.531	-32.67	-0.060	593 (31.1)	569 (29.8)	747 (39.1)

Table 5-6: Distributions of y_{it} to population estimates

(1) Net income: Large enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007												
	March survey	year-on-year rate (FY 2007)	1220	2.666	-1.386	0.003	480	(39.3)	396	(32.5)	344	(28.2)
	June survey	revision rate (FY 2006)	1221	0.629	-0.276	0.001	573	(46.9)	200	(16.4)	448	(36.7)
		revision rate (FY 2007)	1221	0.738	-0.329	0.004	504	(41.3)	296	(24.2)	421	(34.5)
	Sept. survey	revision rate (FY 2007)	1216	0.521	-2.395	-0.001	263	(21.6)	697	(57.3)	256	(21.1)
	Dec. survey	revision rate (FY 2007)	1211	0.931	-1.574	-0.002	478	(39.5)	270	(22.3)	463	(38.2)
FY 2008												
	March survey	revision rate (FY 2007)	1210	0.911	-0.653	-0.005	274	(22.6)	519	(42.9)	417	(34.5)
		year-on-year rate (FY 2008)	1198	0.690	-1.087	0.004	505	(42.2)	381	(31.8)	312	(26.0)
	June survey	revision rate (FY 2007)	1198	1.476	-0.917	-0.003	500	(41.7)	190	(15.9)	508	(42.4)
		revision rate (FY 2008)	1198	1.413	-1.053	0.000	461	(38.5)	300	(25.0)	437	(36.5)
	Sept. survey	revision rate (FY 2008)	1188	0.883	-1.301	-0.005	187	(15.7)	673	(56.6)	328	(27.6)
	Dec. survey	revision rate (FY 2008)	1185	0.657	-1.272	-0.013	363	(30.6)	264	(22.3)	558	(47.1)
FY 2009												
	March survey	revision rate (FY 2008)	1181	22.15	-21.57	-0.018	213	(18.0)	369	(31.2)	599	(50.7)
		year-on-year rate (FY 2009)	1172	26.84	-26.44	0.003	439	(37.5)	388	(33.1)	345	(29.4)
	June survey	revision rate (FY 2008)	1173	8.357	-2.488	-0.023	425	(36.2)	181	(15.4)	567	(48.3)
		revision rate (FY 2009)	1173	2.405	-4.349	-0.005	442	(37.7)	265	(22.6)	466	(39.7)

(2) Net income: Medium-sized enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007												
	March survey	year-on-year rate (FY 2007)	1684	2.183	-2.266	0.009	857	(50.9)	282	(16.7)	545	(32.4)
	June survey	revision rate (FY 2006)	1686	1.321	-3.085	-0.005	714	(42.3)	369	(21.9)	603	(35.8)
		revision rate (FY 2007)	1686	0.792	-0.596	0.002	556	(33.0)	575	(34.1)	555	(32.9)
	Sept. survey	revision rate (FY 2007)	1671	0.368	-0.825	-0.002	511	(30.6)	588	(35.2)	572	(34.2)
	Dec. survey	revision rate (FY 2007)	1656	1.647	-0.920	0.000	646	(39.0)	288	(17.4)	722	(43.6)
FY 2008												
	March survey	revision rate (FY 2007)	1649	1.164	-0.768	-0.003	508	(30.8)	463	(28.1)	678	(41.1)
		year-on-year rate (FY 2008)	1632	3.885	-1.768	0.011	839	(51.4)	240	(14.7)	553	(33.9)
	June survey	revision rate (FY 2007)	1633	0.427	-2.490	-0.007	680	(41.6)	345	(21.1)	608	(37.2)
		revision rate (FY 2008)	1633	0.758	-0.972	-0.002	473	(29.0)	571	(35.0)	589	(36.1)
	Sept. survey	revision rate (FY 2008)	1620	2.211	-3.590	-0.006	400	(24.7)	540	(33.3)	680	(42.0)
	Dec. survey	revision rate (FY 2008)	1616	2.226	-2.351	-0.010	562	(34.8)	277	(17.1)	777	(48.1)
FY 2009												
	March survey	revision rate (FY 2008)	1608	4.086	-2.759	-0.019	405	(25.2)	356	(22.1)	847	(52.7)
		year-on-year rate (FY 2009)	1584	5.590	-5.509	0.013	720	(45.5)	269	(17.0)	595	(37.6)
	June survey	revision rate (FY 2008)	1585	3.460	-9.088	-0.021	639	(40.3)	333	(21.0)	613	(38.7)
		revision rate (FY 2009)	1585	7.893	-2.200	0.003	517	(32.6)	533	(33.6)	535	(33.8)

(3) Net income: Small enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007												
	March survey	year-on-year rate (FY 2007)	3439	3.869	-1.272	0.005	1478	(43.0)	990	(28.8)	971	(28.2)
	June survey	revision rate (FY 2006)	3440	0.652	-0.442	-0.001	1073	(31.2)	1296	(37.7)	1071	(31.1)
		revision rate (FY 2007)	3440	1.434	-1.785	-0.001	934	(27.2)	1469	(42.7)	1037	(30.1)
	Sept. survey	revision rate (FY 2007)	3403	1.841	-3.604	-0.002	870	(25.6)	1378	(40.5)	1155	(33.9)
	Dec. survey	revision rate (FY 2007)	3353	2.085	-1.028	-0.001	1091	(32.5)	939	(28.0)	1323	(39.5)
FY 2008												
	March survey	revision rate (FY 2007)	3389	0.851	-0.841	-0.002	899	(26.5)	1215	(35.9)	1275	(37.6)
		year-on-year rate (FY 2008)	3341	2.209	-1.547	0.006	1436	(43.0)	957	(28.6)	948	(28.4)
	June survey	revision rate (FY 2007)	3344	0.752	-0.866	-0.001	1065	(31.8)	1195	(35.7)	1084	(32.4)
		revision rate (FY 2008)	3344	0.957	-1.330	-0.002	869	(26.0)	1340	(40.1)	1135	(33.9)
	Sept. survey	revision rate (FY 2008)	3299	1.000	-0.606	-0.002	790	(23.9)	1278	(38.7)	1231	(37.3)
	Dec. survey	revision rate (FY 2008)	3257	0.863	-1.879	-0.007	950	(29.2)	856	(26.3)	1451	(44.6)
FY 2009												
	March survey	revision rate (FY 2008)	3316	1.609	-2.197	-0.009	790	(23.8)	1051	(31.7)	1475	(44.5)
		year-on-year rate (FY 2009)	3268	8.076	-1.607	0.016	1285	(39.3)	839	(25.7)	1144	(35.0)
	June survey	revision rate (FY 2008)	3270	4.210	-4.453	-0.001	1082	(33.1)	1135	(34.7)	1053	(32.2)
		revision rate (FY 2009)	3270	0.782	-2.994	-0.002	994	(30.4)	1218	(37.2)	1058	(32.4)

Table 5-7: Distributions of y_{it} to population estimates

(1) Fixed investment: Large enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1244	0.605	-0.566	0.002	474 (38.1)	363 (29.2)	407 (32.7)
	June survey	revision rate (FY 2006)	1245	0.378	-0.327	-0.003	400 (32.1)	225 (18.1)	620 (49.8)
		revision rate (FY 2007)	1245	0.464	-0.728	0.004	482 (38.7)	385 (30.9)	378 (30.4)
	Sept. survey	revision rate (FY 2007)	1239	0.460	-0.151	0.001	199 (16.1)	806 (65.1)	234 (18.9)
	Dec. survey	revision rate (FY 2007)	1235	0.458	-0.185	-0.001	422 (34.2)	275 (22.3)	538 (43.6)
FY 2008									
	March survey	revision rate (FY 2007)	1233	0.410	-0.176	-0.001	206 (16.7)	699 (56.7)	328 (26.6)
		year-on-year rate (FY 2008)	1223	0.341	-1.173	-0.003	428 (35.0)	378 (30.9)	417 (34.1)
	June survey	revision rate (FY 2007)	1225	0.321	-0.699	-0.004	365 (29.8)	225 (18.4)	635 (51.8)
		revision rate (FY 2008)	1225	0.429	-0.460	0.004	474 (38.7)	386 (31.5)	365 (29.8)
	Sept. survey	revision rate (FY 2008)	1218	0.381	-0.379	-0.001	175 (14.4)	769 (63.1)	274 (22.5)
	Dec. survey	revision rate (FY 2008)	1214	0.183	-0.321	-0.003	352 (29.0)	272 (22.4)	590 (48.6)
FY 2009									
	March survey	revision rate (FY 2008)							
		year-on-year rate (FY 2009)	1184	0.298	-1.115	-0.011	244 (20.6)	360 (30.4)	580 (49.0)
	June survey	revision rate (FY 2008)	1201	0.320	-0.380	-0.003	443 (36.9)	193 (16.1)	565 (47.0)
		revision rate (FY 2009)	1201	0.536	-2.130	-0.014	325 (27.1)	347 (28.9)	529 (44.0)

(2) Fixed investment: Medium-sized enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1234	3.721	-0.831	0.001	497 (40.3)	232 (18.8)	505 (40.9)
	June survey	revision rate (FY 2006)	1235	1.192	-0.827	0.001	391 (31.7)	446 (36.1)	398 (32.2)
		revision rate (FY 2007)	1235	0.327	-0.697	0.003	288 (23.3)	729 (59.0)	218 (17.7)
	Sept. survey	revision rate (FY 2007)	1227	0.263	-1.047	-0.001	291 (23.7)	634 (51.7)	302 (24.6)
	Dec. survey	revision rate (FY 2007)	1218	0.410	-0.389	0.000	418 (34.3)	387 (31.8)	413 (33.9)
FY 2008									
	March survey	revision rate (FY 2007)	1219	0.362	-0.479	-0.001	293 (24.0)	577 (47.3)	349 (28.6)
		year-on-year rate (FY 2008)	1200	0.752	-3.437	-0.003	457 (38.1)	232 (19.3)	511 (42.6)
	June survey	revision rate (FY 2007)	1202	0.329	-0.584	-0.001	361 (30.0)	444 (36.9)	397 (33.0)
		revision rate (FY 2008)	1202	0.612	-0.391	0.003	275 (22.9)	699 (58.2)	228 (19.0)
	Sept. survey	revision rate (FY 2008)	1197	0.491	-0.569	0.000	263 (22.0)	622 (52.0)	312 (26.1)
	Dec. survey	revision rate (FY 2008)	1192	0.604	-0.714	-0.003	349 (29.3)	366 (30.7)	477 (40.0)
FY 2009									
	March survey	revision rate (FY 2008)							
		year-on-year rate (FY 2009)	1167	0.666	-1.003	-0.027	270 (23.1)	205 (17.6)	692 (59.3)
	June survey	revision rate (FY 2008)	1172	1.331	-0.363	0.002	382 (32.6)	442 (37.7)	348 (29.7)
		revision rate (FY 2009)	1172	1.108	-0.473	0.002	254 (21.7)	650 (55.5)	268 (22.9)

(3) Fixed investment: Small enterprises of manufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	2030	0.758	-2.050	-0.008	519 (25.6)	740 (36.5)	771 (38.0)
	June survey	revision rate (FY 2006)	2031	11.49	-1.811	0.006	430 (21.2)	1261 (62.1)	340 (16.7)
		revision rate (FY 2007)	2031	1.015	-0.700	0.004	398 (19.6)	1372 (67.6)	261 (12.9)
	Sept. survey	revision rate (FY 2007)	2014	1.760	-0.490	0.003	423 (21.0)	1261 (62.6)	330 (16.4)
	Dec. survey	revision rate (FY 2007)	2002	0.760	-0.422	0.004	532 (26.6)	1067 (53.3)	403 (20.1)
FY 2008									
	March survey	revision rate (FY 2007)	2004	0.995	-1.984	0.001	421 (21.0)	1195 (59.6)	388 (19.4)
		year-on-year rate (FY 2008)	1977	1.725	-2.779	-0.011	452 (22.9)	762 (38.5)	763 (38.6)
	June survey	revision rate (FY 2007)	1977	0.698	-0.703	0.001	383 (19.4)	1235 (62.5)	359 (18.2)
		revision rate (FY 2008)	1977	0.945	-0.887	0.004	398 (20.1)	1304 (66.0)	275 (13.9)
	Sept. survey	revision rate (FY 2008)	1957	9.080	-0.546	0.006	379 (19.4)	1212 (61.9)	366 (18.7)
	Dec. survey	revision rate (FY 2008)	1945	1.188	-0.810	0.000	432 (22.2)	1040 (53.5)	473 (24.3)
FY 2009									
	March survey	revision rate (FY 2008)							
		year-on-year rate (FY 2009)	1885	1.656	-2.144	-0.022	266 (14.1)	744 (39.5)	875 (46.4)
	June survey	revision rate (FY 2008)	1896	0.412	-0.410	0.001	358 (18.9)	1224 (64.6)	314 (16.6)
		revision rate (FY 2009)	1896	1.134	-0.711	0.002	318 (16.8)	1320 (69.6)	258 (13.6)

Table 5-8: Distributions of y_{it} to population estimates

(1) Fixed investment: Large enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1216	2.293	-1.830	0.003	329 (27.1)	472 (38.8)	415 (34.1)
	June survey	revision rate (FY 2006)	1218	0.569	-0.290	0.000	415 (34.1)	370 (30.4)	433 (35.6)
		revision rate (FY 2007)	1218	1.000	-1.722	0.002	403 (33.1)	524 (43.0)	291 (23.9)
	Sept. survey	revision rate (FY 2007)	1212	0.566	-0.318	0.001	207 (17.1)	822 (67.8)	183 (15.1)
	Dec. survey	revision rate (FY 2007)	1208	1.459	-0.370	0.003	416 (34.4)	383 (31.7)	409 (33.9)
FY 2008									
	March survey	revision rate (FY 2007)	1207	0.341	-1.644	-0.004	252 (20.9)	689 (57.1)	266 (22.0)
		year-on-year rate (FY 2008)	1194	0.678	-0.569	-0.001	331 (27.7)	448 (37.5)	415 (34.8)
	June survey	revision rate (FY 2007)	1195	0.702	-1.060	0.000	424 (35.5)	352 (29.5)	419 (35.1)
		revision rate (FY 2008)	1195	1.325	-0.907	0.000	383 (32.1)	514 (43.0)	298 (24.9)
	Sept. survey	revision rate (FY 2008)	1185	0.466	-0.743	0.000	177 (14.9)	805 (67.9)	203 (17.1)
	Dec. survey	revision rate (FY 2008)	1182	0.933	-1.176	-0.001	389 (32.9)	377 (31.9)	416 (35.2)
FY 2009									
	March survey	revision rate (FY 2008)	[redacted]						
		year-on-year rate (FY 2009)	1139	0.822	-0.768	-0.002	244 (21.4)	438 (38.5)	457 (40.1)
	June survey	revision rate (FY 2008)	1157	1.480	-1.521	-0.003	419 (36.2)	356 (30.8)	382 (33.0)
		revision rate (FY 2009)	1157	0.430	-0.604	-0.001	355 (30.7)	461 (39.8)	341 (29.5)

(2) Fixed investment: Medium-sized enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1687	6.880	-1.751	0.002	443 (26.3)	645 (38.2)	599 (35.5)
	June survey	revision rate (FY 2006)	1689	0.496	-1.342	0.001	468 (27.7)	895 (53.0)	326 (19.3)
		revision rate (FY 2007)	1689	0.696	-1.464	0.001	350 (20.7)	1116 (66.1)	223 (13.2)
	Sept. survey	revision rate (FY 2007)	1672	1.319	-0.941	0.001	305 (18.2)	1099 (65.7)	268 (16.0)
	Dec. survey	revision rate (FY 2007)	1657	0.271	-1.019	-0.001	436 (26.3)	824 (49.7)	397 (24.0)
FY 2008									
	March survey	revision rate (FY 2007)	1649	0.390	-3.374	-0.003	308 (18.7)	985 (59.7)	356 (21.6)
		year-on-year rate (FY 2008)	1633	2.790	-1.332	-0.001	415 (25.4)	610 (37.4)	608 (37.2)
	June survey	revision rate (FY 2007)	1633	0.747	-0.359	0.001	433 (26.5)	885 (54.2)	315 (19.3)
		revision rate (FY 2008)	1633	1.685	-0.305	0.003	297 (18.2)	1107 (67.8)	229 (14.0)
	Sept. survey	revision rate (FY 2008)	1620	1.740	-4.128	-0.003	265 (16.4)	1066 (65.8)	289 (17.8)
	Dec. survey	revision rate (FY 2008)	1616	1.592	-2.770	-0.001	405 (25.1)	811 (50.2)	400 (24.8)
FY 2009									
	March survey	revision rate (FY 2008)	[redacted]						
		year-on-year rate (FY 2009)	1550	0.823	-8.159	-0.021	296 (19.1)	581 (37.5)	673 (43.4)
	June survey	revision rate (FY 2008)	1562	4.990	-0.764	0.005	449 (28.7)	827 (52.9)	286 (18.3)
		revision rate (FY 2009)	1562	2.503	-0.982	0.003	286 (18.3)	1020 (65.3)	256 (16.4)

(3) Fixed investment: Small enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	3447	5.319	-4.386	-0.005	433 (12.6)	2116 (61.4)	898 (26.1)
	June survey	revision rate (FY 2006)	3447	1.773	-0.515	0.001	441 (12.8)	2670 (77.5)	336 (9.7)
		revision rate (FY 2007)	3447	1.877	-4.343	0.002	445 (12.9)	2739 (79.5)	263 (7.6)
	Sept. survey	revision rate (FY 2007)	3408	3.506	-0.410	0.002	454 (13.3)	2657 (78.0)	297 (8.7)
	Dec. survey	revision rate (FY 2007)	3359	2.856	-0.206	0.002	530 (15.8)	2443 (72.7)	386 (11.5)
FY 2008									
	March survey	revision rate (FY 2007)	3400	1.110	-0.342	0.001	461 (13.6)	2505 (73.7)	434 (12.8)
		year-on-year rate (FY 2008)	3353	0.651	-3.141	-0.008	420 (12.5)	2082 (62.1)	851 (25.4)
	June survey	revision rate (FY 2007)	3357	0.808	-0.459	0.001	474 (14.1)	2559 (76.2)	324 (9.7)
		revision rate (FY 2008)	3357	1.150	-0.426	0.003	463 (13.8)	2626 (78.2)	268 (8.0)
	Sept. survey	revision rate (FY 2008)	3310	2.826	-0.501	0.002	442 (13.4)	2557 (77.3)	311 (9.4)
	Dec. survey	revision rate (FY 2008)	3265	0.523	-0.562	0.001	492 (15.1)	2385 (73.0)	388 (11.9)
FY 2009									
	March survey	revision rate (FY 2008)	[redacted]						
		year-on-year rate (FY 2009)	3236	2.909	-3.145	-0.014	281 (8.7)	2001 (61.8)	954 (29.5)
	June survey	revision rate (FY 2008)	3245	2.393	-1.281	0.002	423 (13.0)	2548 (78.5)	274 (8.4)
		revision rate (FY 2009)	3245	0.671	-2.254	0.001	395 (12.2)	2620 (80.7)	230 (7.1)

Table 5-9: Distributions of y_{it} to population estimates

(1) Software investment: Large enterprises of manufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007											
March survey	year-on-year rate (FY 2007)	1231	1.282	-1.050	-0.003	267	(21.7)	625	(50.8)	339	(27.5)
June survey	revision rate (FY 2006)	1234	2.213	-0.875	0.002	351	(28.4)	541	(43.8)	342	(27.7)
	revision rate (FY 2007)	1234	0.668	-0.906	0.002	291	(23.6)	715	(57.9)	228	(18.5)
Sept. survey	revision rate (FY 2007)	1228	0.732	-0.514	0.002	152	(12.4)	954	(77.7)	122	(9.9)
Dec. survey	revision rate (FY 2007)	1224	1.717	-0.199	0.007	362	(29.6)	536	(43.8)	326	(26.6)
FY 2008											
March survey	revision rate (FY 2007)	1222	0.839	-0.820	0.000	163	(13.3)	871	(71.3)	188	(15.4)
	year-on-year rate (FY 2008)	1211	2.066	-1.153	-0.001	273	(22.5)	623	(51.4)	315	(26.0)
June survey	revision rate (FY 2007)	1214	0.715	-0.573	0.000	381	(31.4)	504	(41.5)	329	(27.1)
	revision rate (FY 2008)	1214	1.855	-0.672	0.007	301	(24.8)	679	(55.9)	234	(19.3)
Sept. survey	revision rate (FY 2008)	1209	0.785	-0.517	0.000	144	(11.9)	924	(76.4)	141	(11.7)
Dec. survey	revision rate (FY 2008)	1205	1.237	-0.580	0.000	334	(27.7)	522	(43.3)	349	(29.0)
FY 2009											
March survey	revision rate (FY 2008)										
	year-on-year rate (FY 2009)	1177	0.296	-2.228	-0.010	192	(16.3)	559	(47.5)	426	(36.2)
June survey	revision rate (FY 2008)	1192	0.942	-0.897	0.001	429	(36.0)	435	(36.5)	328	(27.5)
	revision rate (FY 2009)	1192	0.988	-1.499	-0.004	280	(23.5)	602	(50.5)	310	(26.0)

(2) Software investment: Medium-sized enterprises of manufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007											
March survey	year-on-year rate (FY 2007)	1235	4.456	-5.113	-0.006	222	(18.0)	770	(62.3)	243	(19.7)
June survey	revision rate (FY 2006)	1235	1.414	-0.914	0.000	164	(13.3)	936	(75.8)	135	(10.9)
	revision rate (FY 2007)	1235	1.404	-0.928	0.007	118	(9.6)	1046	(84.7)	71	(5.7)
Sept. survey	revision rate (FY 2007)	1227	0.451	-1.228	-0.005	134	(10.9)	993	(80.9)	100	(8.1)
Dec. survey	revision rate (FY 2007)	1218	11.80	-0.911	0.010	176	(14.4)	877	(72.0)	165	(13.5)
FY 2008											
March survey	revision rate (FY 2007)	1219	0.633	-1.736	-0.005	105	(8.6)	978	(80.2)	136	(11.2)
	year-on-year rate (FY 2008)	1200	4.776	-11.27	0.009	219	(18.3)	753	(62.8)	228	(19.0)
June survey	revision rate (FY 2007)	1201	0.712	-0.972	-0.001	173	(14.4)	883	(73.5)	145	(12.1)
	revision rate (FY 2008)	1201	0.921	-1.172	-0.001	86	(7.2)	1036	(86.3)	79	(6.6)
Sept. survey	revision rate (FY 2008)	1196	0.808	-1.042	-0.001	116	(9.7)	980	(81.9)	100	(8.4)
Dec. survey	revision rate (FY 2008)	1191	0.703	-3.497	-0.008	157	(13.2)	862	(72.4)	172	(14.4)
FY 2009											
March survey	revision rate (FY 2008)										
	year-on-year rate (FY 2009)	1166	1.180	-6.628	-0.019	147	(12.6)	712	(61.1)	307	(26.3)
June survey	revision rate (FY 2008)	1171	2.454	-2.025	0.000	159	(13.6)	863	(73.7)	149	(12.7)
	revision rate (FY 2009)	1171	15.86	-2.470	0.017	113	(9.6)	982	(83.9)	76	(6.5)

(3) Software investment: Small enterprises of manufacturing

		Number of data	Max	Min	Average	ratio of positive data(%)		ratio of zero(%)		ratio of negative data(%)	
FY 2007											
March survey	year-on-year rate (FY 2007)	2030	2.395	-2.732	-0.002	171	(8.4)	1620	(79.8)	239	(11.8)
June survey	revision rate (FY 2006)	2031	1.163	-2.732	-0.003	126	(6.2)	1815	(89.4)	90	(4.4)
	revision rate (FY 2007)	2031	1.780	-2.657	0.003	140	(6.9)	1820	(89.6)	71	(3.5)
Sept. survey	revision rate (FY 2007)	2014	1.818	-1.221	0.001	128	(6.4)	1776	(88.2)	110	(5.5)
Dec. survey	revision rate (FY 2007)	2002	1.685	-1.293	0.001	141	(7.0)	1746	(87.2)	115	(5.7)
FY 2008											
March survey	revision rate (FY 2007)	2004	1.107	-1.687	-0.004	106	(5.3)	1769	(88.3)	129	(6.4)
	year-on-year rate (FY 2008)	1977	4.260	-2.099	-0.001	172	(8.7)	1573	(79.6)	232	(11.7)
June survey	revision rate (FY 2007)	1977	2.197	-1.316	-0.001	104	(5.3)	1788	(90.4)	85	(4.3)
	revision rate (FY 2008)	1977	7.383	-1.734	0.009	130	(6.6)	1776	(89.8)	71	(3.6)
Sept. survey	revision rate (FY 2008)	1957	2.891	-6.184	0.000	114	(5.8)	1733	(88.6)	110	(5.6)
Dec. survey	revision rate (FY 2008)	1945	1.774	-1.663	-0.002	131	(6.7)	1694	(87.1)	120	(6.2)
FY 2009											
March survey	revision rate (FY 2008)										
	year-on-year rate (FY 2009)	1885	4.830	-3.985	-0.016	88	(4.7)	1493	(79.2)	304	(16.1)
June survey	revision rate (FY 2008)	1896	3.974	-1.993	0.000	89	(4.7)	1709	(90.1)	98	(5.2)
	revision rate (FY 2009)	1896	1.757	-4.532	0.004	99	(5.2)	1723	(90.9)	74	(3.9)

Table 5-10: Distributions of y_{it} to population estimates

(1) Software investment: Large enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1203	0.922	-3.014	-0.004	235 (19.5)	694 (57.7)	274 (22.8)
	June survey	revision rate (FY 2006)	1205	1.075	-0.806	0.000	290 (24.1)	617 (51.2)	298 (24.7)
		revision rate (FY 2007)	1205	0.626	-1.701	-0.001	245 (20.3)	744 (61.7)	216 (17.9)
	Sept. survey	revision rate (FY 2007)	1199	0.448	-0.553	0.000	140 (11.7)	944 (78.7)	115 (9.6)
	Dec. survey	revision rate (FY 2007)	1194	0.759	-0.741	-0.002	285 (23.9)	638 (53.4)	271 (22.7)
FY 2008									
	March survey	revision rate (FY 2007)	1195	1.867	-0.196	0.002	161 (13.5)	867 (72.6)	167 (14.0)
		year-on-year rate (FY 2008)	1183	1.833	-0.639	0.003	231 (19.5)	659 (55.7)	293 (24.8)
	June survey	revision rate (FY 2007)	1184	3.079	-1.368	0.006	326 (27.5)	598 (50.5)	260 (22.0)
		revision rate (FY 2008)	1184	1.932	-0.804	0.007	273 (23.1)	739 (62.4)	172 (14.5)
	Sept. survey	revision rate (FY 2008)	1175	0.464	-0.718	0.000	132 (11.2)	916 (78.0)	127 (10.8)
	Dec. survey	revision rate (FY 2008)	1172	1.170	-2.810	-0.008	299 (25.5)	617 (52.6)	256 (21.8)
FY 2009									
	March survey	revision rate (FY 2008)							
		year-on-year rate (FY 2009)	1132	0.478	-0.699	-0.005	182 (16.1)	620 (54.8)	330 (29.2)
	June survey	revision rate (FY 2008)	1149	4.683	-0.918	0.006	307 (26.7)	562 (48.9)	280 (24.4)
		revision rate (FY 2009)	1149	3.774	-1.626	0.003	226 (19.7)	675 (58.7)	248 (21.6)

(2) Software investment: Medium-sized enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	1687	9.260	-2.052	0.010	227 (13.5)	1139 (67.5)	321 (19.0)
	June survey	revision rate (FY 2006)	1689	1.597	-5.481	0.001	222 (13.1)	1303 (77.1)	164 (9.7)
		revision rate (FY 2007)	1689	1.555	-11.00	-0.004	165 (9.8)	1416 (83.8)	108 (6.4)
	Sept. survey	revision rate (FY 2007)	1672	1.348	-0.761	-0.001	153 (9.2)	1403 (83.9)	116 (6.9)
	Dec. survey	revision rate (FY 2007)	1657	2.058	-1.081	0.001	196 (11.8)	1269 (76.6)	192 (11.6)
FY 2008									
	March survey	revision rate (FY 2007)	1649	0.898	-1.576	-0.001	153 (9.3)	1329 (80.6)	167 (10.1)
		year-on-year rate (FY 2008)	1634	2.136	-2.973	-0.002	240 (14.7)	1079 (66.0)	315 (19.3)
	June survey	revision rate (FY 2007)	1634	3.344	-1.814	0.001	205 (12.5)	1248 (76.4)	181 (11.1)
		revision rate (FY 2008)	1634	1.987	-1.279	0.002	144 (8.8)	1382 (84.6)	108 (6.6)
	Sept. survey	revision rate (FY 2008)	1620	0.531	-0.264	0.002	147 (9.1)	1350 (83.3)	123 (7.6)
	Dec. survey	revision rate (FY 2008)	1616	0.961	-0.659	0.000	242 (15.0)	1175 (72.7)	199 (12.3)
FY 2009									
	March survey	revision rate (FY 2008)							
		year-on-year rate (FY 2009)	1551	3.479	-3.325	-0.008	168 (10.8)	1022 (65.9)	361 (23.3)
	June survey	revision rate (FY 2008)	1562	2.565	-0.707	0.005	207 (13.3)	1185 (75.9)	170 (10.9)
		revision rate (FY 2009)	1562	1.351	-0.951	0.002	132 (8.5)	1302 (83.4)	128 (8.2)

(3) Software investment: Small enterprises of nonmanufacturing

			Number of data	Max	Min	Average	ratio of positive data(%)	ratio of zero(%)	ratio of negative data(%)
FY 2007									
	March survey	year-on-year rate (FY 2007)	3447	3.649	-1.275	-0.002	194 (5.6)	2975 (86.3)	278 (8.1)
	June survey	revision rate (FY 2006)	3447	1.313	-0.681	0.000	144 (4.2)	3165 (91.8)	138 (4.0)
		revision rate (FY 2007)	3447	1.174	-2.772	0.002	175 (5.1)	3190 (92.5)	82 (2.4)
	Sept. survey	revision rate (FY 2007)	3408	0.974	-0.704	0.002	177 (5.2)	3111 (91.3)	120 (3.5)
	Dec. survey	revision rate (FY 2007)	3359	4.799	-1.509	0.001	168 (5.0)	3055 (90.9)	136 (4.0)
FY 2008									
	March survey	revision rate (FY 2007)	3400	1.440	-1.692	-0.002	146 (4.3)	3037 (89.3)	217 (6.4)
		year-on-year rate (FY 2008)	3353	1.381	-5.376	-0.004	171 (5.1)	2908 (86.7)	274 (8.2)
	June survey	revision rate (FY 2007)	3357	2.085	-1.171	0.000	147 (4.4)	3082 (91.8)	128 (3.8)
		revision rate (FY 2008)	3357	3.007	-0.614	0.006	168 (5.0)	3119 (92.9)	70 (2.1)
	Sept. survey	revision rate (FY 2008)	3311	2.478	-1.684	0.001	159 (4.8)	3035 (91.7)	117 (3.5)
	Dec. survey	revision rate (FY 2008)	3265	2.041	-1.911	0.001	142 (4.3)	2997 (91.8)	126 (3.9)
FY 2009									
	March survey	revision rate (FY 2008)							
		year-on-year rate (FY 2009)	3235	4.658	-3.275	-0.011	143 (4.4)	2760 (85.3)	332 (10.3)
	June survey	revision rate (FY 2008)	3244	2.342	-3.156	0.000	129 (4.0)	3008 (92.7)	107 (3.3)
		revision rate (FY 2009)	3244	4.876	-3.276	0.008	155 (4.8)	3021 (93.1)	68 (2.1)

Table 6: Detected outlier in case of range edit with revision rate

June survey in 2009, Sales, revision rate (2009)

Ratio of outlier in data set

(1)Number of data which exceed upper boundary, T_U

		T_U=2	T_U=3	T_U=4
Manufacturing				
	Large	0.50%	0.17%	0.08%
	Medium-sized	0.59%	0.42%	0.34%
	Small	0.84%	0.63%	0.42%
Nonmanufacturing				
	Large	0.34%	0.17%	0.09%
	Medium-sized	0.82%	0.63%	0.38%
	Small	0.91%	0.58%	0.43%

(2)Number of data which fall short of lower boundary, T_L

		T_L=0.3	T_L=0.2	T_L=0.1
Manufacturing				
	Large	0.17%	0.08%	0.08%
	Medium-sized	0.00%	0.00%	0.00%
	Small	0.26%	0.21%	0.05%
Nonmanufacturing				
	Large	0.60%	0.43%	0.34%
	Medium-sized	0.25%	0.19%	0.06%
	Small	0.21%	0.06%	0.00%

Note: revision rate,r_{it}, is calculated by data of the current survey and data of the previous survey.

Table 7: Detected outlier in case of Hidiroglou-Berthelot method

June survey in 2009, Sales, revision rate (2009)

(1) Distribution of E_{it}

		Number of data	Min	First quartile	Median	Third quartile	Max
Manufacturing							
	Large	1,208	-159	-0.86	4.65E-05	0.13	192
	Medium-sized	1,178	-10	-0.27	0.00	0.00	48
	Small	1,915	-32	-0.31	0.00	0.00	70
Nonmanufacturing							
	Large	1,165	-39	-0.40	0.00	2.99E-03	89
	Medium-sized	1,585	-61	-0.13	0.00	0.00	26535
	Small	3,283	-25	-0.14	0.00	4.12E-03	141

(2) D_L and $D_{U,}$ tolerable range

		indicator of variance		tolerable range					
				c=20		c=30		c=40	
		D_L	D_U	Lower	Upper	Lower	Upper	Lower	Upper
Manufacturing									
	Large	0.86	0.13	-17.15	2.68	-25.72	4.02	-34.30	5.37
	Medium-sized	0.27	0.00	-5.33	0.00	-8.00	0.00	-10.66	0.00
	Small	0.31	0.00	-6.11	0.00	-9.17	0.00	-12.23	0.00
Nonmanufacturing									
	Large	0.40	2.99E-03	-8.00	0.06	-12.00	0.09	-16.00	0.12
	Medium-sized	0.13	0.00	-2.63	0.00	-3.94	0.00	-5.25	0.00
	Small	0.14	4.12E-03	-2.87	0.08	-4.31	0.12	-5.75	0.16

(3) Ratio of outlier in data set

		c=20		c=30		c=40	
		Lower	Upper	Lower	Upper	Lower	Upper
Manufacturing							
	Large	0.08%	1.82%	0.08%	1.16%	0.08%	0.66%
	Medium-sized	0.76%	22.75%	0.08%	22.75%	0.00%	22.75%
	Small	0.57%	23.24%	0.37%	23.24%	0.21%	23.24%
Nonmanufacturing							
	Large	1.12%	20.60%	0.52%	18.80%	0.26%	17.51%
	Medium-sized	1.32%	24.10%	0.76%	24.10%	0.63%	24.10%
	Small	0.91%	20.07%	0.58%	17.85%	0.37%	15.32%

Table 8-1: Unit of Distance, D_j

(1) Sales

			Manufacturing			Nonmanufacturing		
			Large	Medium-sized	Small	Large	Medium-sized	Small
FY 2007								
	March survey	year-on-year rate (FY 2007)	0.044	0.066	0.042	0.051	0.050	0.033
	June survey	revision rate (FY 2006)	0.021	0.023	0.024	0.027	0.025	0.016
		revision rate (FY 2007)	0.057	0.049	0.034	0.051	0.035	0.020
	Sept. survey	revision rate (FY 2007)	0.032	0.049	0.027	0.024	0.039	0.026
	Dec. survey	revision rate (FY 2007)	0.039	0.026	0.022	0.047	0.031	0.020
FY 2008								
	March survey	revision rate (FY 2007)	0.025	0.034	0.037	0.025	0.029	0.028
		year-on-year rate (FY 2008)	0.034	0.078	0.045	0.037	0.072	0.031
	June survey	revision rate (FY 2007)	0.025	0.027	0.020	0.047	0.023	0.017
		revision rate (FY 2008)	0.081	0.045	0.042	0.090	0.040	0.024
	Sept. survey	revision rate (FY 2008)	0.039	0.059	0.038	0.045	0.038	0.024
	Dec. survey	revision rate (FY 2008)	0.064	0.043	0.035	0.081	0.026	0.023
FY 2009								
	March survey	revision rate (FY 2008)	0.131	0.092	0.045	0.088	0.050	0.035
		year-on-year rate (FY 2009)	0.099	0.128	0.083	0.154	0.099	0.048
	June survey	revision rate (FY 2008)	0.046	0.036	0.022	0.053	0.030	0.018
		revision rate (FY 2009)	0.237	0.071	0.065	0.171	0.053	0.034

(2)Current profits

			Manufacturing			Nonmanufacturing		
			Large	Medium-sized	Small	Large	Medium-sized	Small
FY 2007								
	March survey	year-on-year rate (FY 2007)	0.142	0.264	0.259	0.155	0.275	0.172
	June survey	revision rate (FY 2006)	0.092	0.129	0.169	0.106	0.170	0.120
		revision rate (FY 2007)	0.209	0.155	0.172	0.185	0.199	0.127
	Sept. survey	revision rate (FY 2007)	0.104	0.191	0.183	0.096	0.135	0.128
	Dec. survey	revision rate (FY 2007)	0.138	0.155	0.185	0.219	0.169	0.132
FY 2008								
	March survey	revision rate (FY 2007)	0.096	0.162	0.180	0.156	0.153	0.117
		year-on-year rate (FY 2008)	0.191	0.278	0.314	0.151	0.339	0.207
	June survey	revision rate (FY 2007)	0.182	0.171	0.195	0.181	0.156	0.136
		revision rate (FY 2008)	0.329	0.218	0.222	0.218	0.157	0.136
	Sept. survey	revision rate (FY 2008)	0.146	0.229	0.256	0.203	0.254	0.144
	Dec. survey	revision rate (FY 2008)	0.313	0.291	0.284	0.329	0.187	0.183
FY 2009								
	March survey	revision rate (FY 2008)	0.980	0.451	0.348	0.540	0.267	0.195
		year-on-year rate (FY 2009)	1.191	0.953	0.950	0.326	0.593	0.295
	June survey	revision rate (FY 2008)	0.702	0.302	0.492	0.357	0.288	0.177
		revision rate (FY 2009)	1.370	0.809	0.973	0.710	0.280	0.187

(3)Net income

			Manufacturing			Nonmanufacturing		
			Large	Medium-sized	Small	Large	Medium-sized	Small
FY 2007								
	March survey	year-on-year rate (FY 2007)	0.248	0.659	0.496	0.494	0.677	0.313
	June survey	revision rate (FY 2006)	0.278	0.324	0.317	0.246	0.532	0.209
		revision rate (FY 2007)	0.280	0.291	0.256	0.347	0.310	0.165
	Sept. survey	revision rate (FY 2007)	0.145	0.273	0.227	0.167	0.234	0.200
	Dec. survey	revision rate (FY 2007)	0.194	0.246	0.312	0.307	0.370	0.255
FY 2008								
	March survey	revision rate (FY 2007)	0.166	0.265	0.287	0.377	0.326	0.214
		year-on-year rate (FY 2008)	0.278	0.463	0.497	0.522	0.816	0.398
	June survey	revision rate (FY 2007)	0.238	0.359	0.317	0.356	0.419	0.261
		revision rate (FY 2008)	0.378	0.329	0.275	0.374	0.229	0.198
	Sept. survey	revision rate (FY 2008)	0.143	0.358	0.336	0.273	0.362	0.208
	Dec. survey	revision rate (FY 2008)	0.411	0.474	0.422	0.489	0.425	0.334
FY 2009								
	March survey	revision rate (FY 2008)	1.372	0.860	0.647	0.880	0.492	0.398
		year-on-year rate (FY 2009)	343.7	4.946	3.302	0.884	1.246	0.735
	June survey	revision rate (FY 2008)	318.0	4.015	1.880	0.951	0.923	0.451
		revision rate (FY 2009)	4.244	1.656	2.720	1.202	0.495	0.309

Table 8-2: Unit of Distance, D_j

(4)Fixed investment

			Manufacturing			Nonmanufacturing		
			Large	Medium-sized	Small	Large	Medium-sized	Small
FY 2007								
	March survey	year-on-year rate (FY 2007)	0.269	0.519	0.660	0.225	0.413	0.339
	June survey	revision rate (FY 2006)	0.194	0.185	0.109	0.173	0.135	0.072
		revision rate (FY 2007)	0.289	0.222	0.211	0.202	0.150	0.155
	Sept. survey	revision rate (FY 2007)	0.069	0.151	0.158	0.048	0.113	0.105
	Dec. survey	revision rate (FY 2007)	0.132	0.194	0.219	0.133	0.158	0.114
FY 2008								
	March survey	revision rate (FY 2007)	0.064	0.197	0.190	0.164	0.162	0.104
		year-on-year rate (FY 2008)	0.281	0.629	0.684	0.267	0.406	0.398
	June survey	revision rate (FY 2007)	0.173	0.177	0.131	0.179	0.110	0.080
		revision rate (FY 2008)	0.247	0.273	0.277	0.264	0.157	0.151
	Sept. survey	revision rate (FY 2008)	0.084	0.203	0.194	0.069	0.116	0.136
	Dec. survey	revision rate (FY 2008)	0.138	0.244	0.156	0.191	0.144	0.103
FY 2009								
	March survey	revision rate (FY 2008)						
		year-on-year rate (FY 2009)	0.293	0.546	0.633	0.353	0.649	0.417
	June survey	revision rate (FY 2008)	0.224	0.177	0.165	0.224	0.145	0.091
		revision rate (FY 2009)	0.452	0.354	0.326	0.288	0.273	0.201

(5)Software investment

			Manufacturing			Nonmanufacturing		
			Large	Medium-sized	Small	Large	Medium-sized	Small
FY 2007								
	March survey	year-on-year rate (FY 2007)	0.371	1.153	0.926	0.274	0.359	0.478
	June survey	revision rate (FY 2006)	0.395	0.410	0.285	0.294	0.188	0.148
		revision rate (FY 2007)	0.498	0.514	0.449	0.266	0.133	0.174
	Sept. survey	revision rate (FY 2007)	0.111	0.394	0.302	0.084	0.132	0.154
	Dec. survey	revision rate (FY 2007)	0.252	0.452	0.413	0.295	0.150	0.169
FY 2008								
	March survey	revision rate (FY 2007)	0.169	0.274	0.375	0.166	0.194	0.141
		year-on-year rate (FY 2008)	0.307	1.069	1.139	0.279	0.389	0.428
	June survey	revision rate (FY 2007)	0.349	0.474	0.354	0.297	0.210	0.134
		revision rate (FY 2008)	0.375	0.397	0.587	0.440	0.201	0.187
	Sept. survey	revision rate (FY 2008)	0.129	0.297	0.349	0.097	0.121	0.182
	Dec. survey	revision rate (FY 2008)	0.229	0.494	0.384	0.291	0.169	0.155
FY 2009								
	March survey	revision rate (FY 2008)						
		year-on-year rate (FY 2009)	0.389	1.191	0.912	0.371	0.680	0.593
	June survey	revision rate (FY 2008)	0.383	0.457	0.283	0.481	0.417	0.189
		revision rate (FY 2009)	0.505	0.743	0.549	0.629	0.306	0.306

Table 9: Standard “C” and number of detected outliers

	C=25	C=50	C=75	C=100	C=125
Number of detected outliers	**14**	**2**	**2**	**1**	**0**
Sales	**0**	**0**	**0**	**0**	**0**
manufacturing					
Large	0	0	0	0	0
Medium-sized	0	0	0	0	0
Small	0	0	0	0	0
nonmanufacturing					
Large	0	0	0	0	0
Medium-sized	0	0	0	0	0
Small	0	0	0	0	0
Current profits	**0**	**0**	**0**	**0**	**0**
manufacturing					
Large	0	0	0	0	0
Medium-sized	0	0	0	0	0
Small	0	0	0	0	0
nonmanufacturing					
Large	0	0	0	0	0
Medium-sized	0	0	0	0	0
Small	0	0	0	0	0
Net income	**2**	**0**	**0**	**0**	**0**
manufacturing					
Large	0	0	0	0	0
Medium-sized	0	0	0	0	0
Small	0	0	0	0	0
nonmanufacturing					
Large	2	0	0	0	0
Medium-sized	0	0	0	0	0
Small	0	0	0	0	0
Fixed investment	**7**	**1**	**1**	**1**	**0**
manufacturing					
Large	0	0	0	0	0
Medium-sized	0	0	0	0	0
Small	2	1	1	1	0
nonmanufacturing					
Large	0	0	0	0	0
Medium-sized	2	0	0	0	0
Small	3	0	0	0	0
Software investment	**5**	**1**	**1**	**0**	**0**
manufacturing					
Large	0	0	0	0	0
Medium-sized	1	0	0	0	0
Small	0	0	0	0	0
nonmanufacturing					
Large	0	0	0	0	0
Medium-sized	3	1	1	0	0
Small	1	0	0	0	0

Table 10: Output of outlier treatment

(1)y_{it} and z_{it} of detected outlier data

	y_{it}	z_{it}
(a)June survey in 2007: Small enterprises of manufacturing: Fixed investment(2006)	+11.49	+104.81
(b)June survey in 2007: Medium-sized enterprises of nonmanufacturing: Software investment(2007)	▲11.00	+82.18
(c) September survey in 2008: Small enterprises of manufacturing: Fixed investment(2008)	+9.08	+46.11

(2)y_{it} after outlier treatment and population estimates

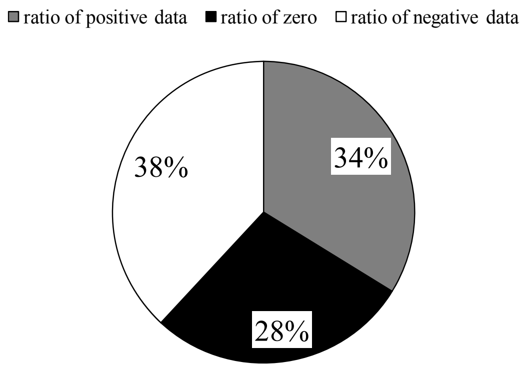
		Original data	After outlier treatment		
			(A) Cold Deck Imputation	(B) Cell Mean Imputation	(C) Growth rate Imputation
(a)June survey in 2007: Small enterprises of manufacturing: Fixed investment(2006)	Revision rate	+11.51	+0.02	▲0.87	▲0.05
	y_{it} of outlier	+11.49	0.00	▲0.89	▲0.07
(b)June survey in 2007: Medium-sized enterprises of nonmanufacturing: Software investment(2007)	Revision rate	▲6.32	+4.68	▲23.15	+12.06
	y_{it} of outlier	▲11.00	0.00	▲27.83	+7.38
(c)September survey in 2008: Small enterprises of manufacturing: Fixed investment(2008)	Revision rate	+11.34	+2.16	+0.20	+2.20
	y_{it} of outlier	+9.08	0.00	▲1.94	+0.04

Figure 1-1: Distributions of y_{it} to population estimates
: Average of ten surveys

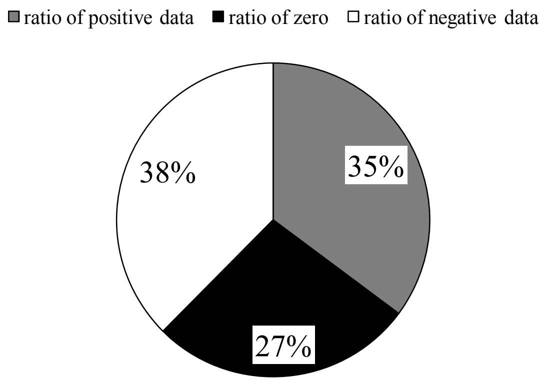
Note: Percentage of zero of “Fixed investment” and “Software investment” is much higher than those of “Sales” and “Profits”.

Large enterprises of manufacturing

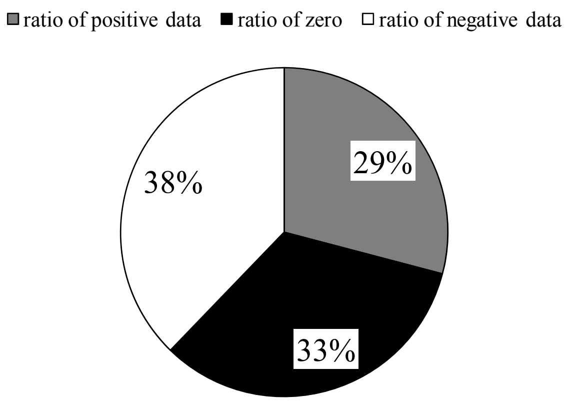
(1) Sales



(2) Current profits



(3) Fixed investment



(4) Software investment

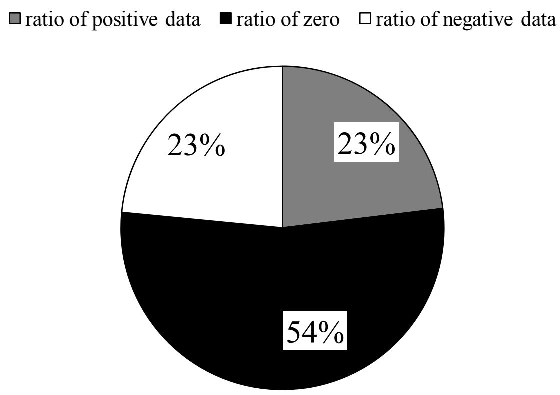
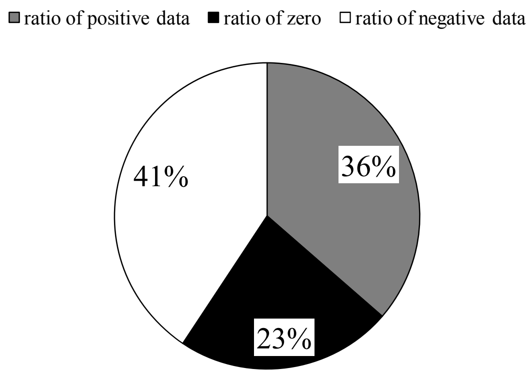


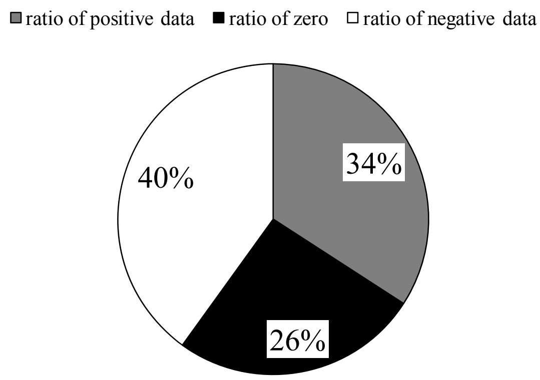
Figure 1-2: Distributions of y_{it} to population estimates (2)
: Average of ten surveys

Small enterprises of manufacturing

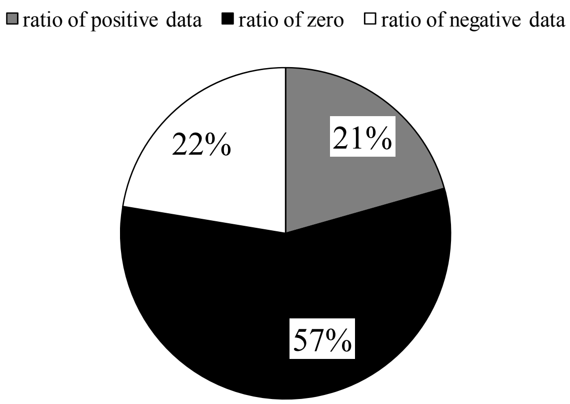
(1) Sales



(2) Current profits



(3) Fixed investment



(4) Software investment

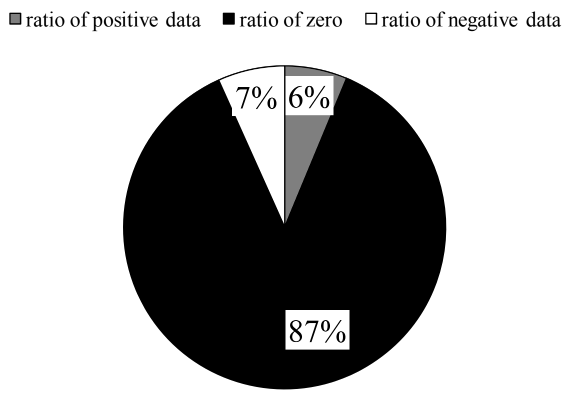
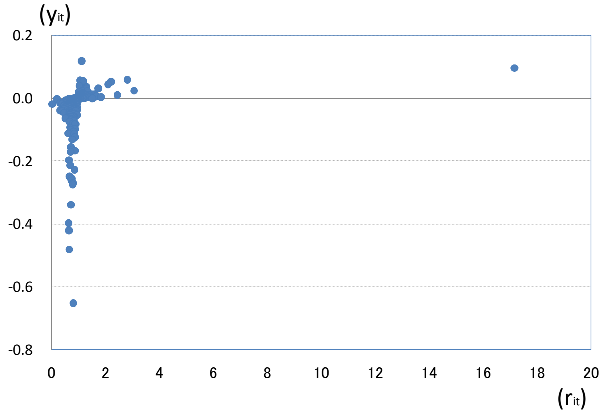
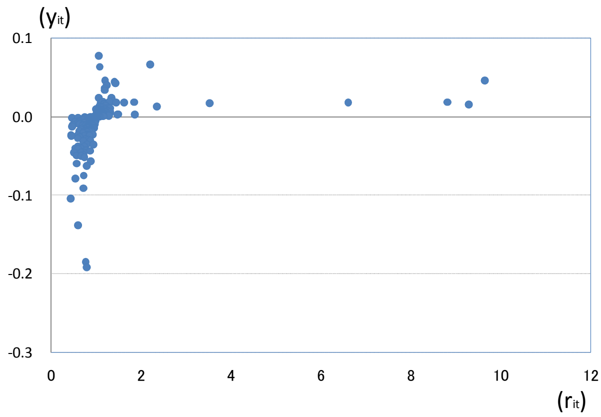


Figure 2-1: Relationship between r_i and y_{it}

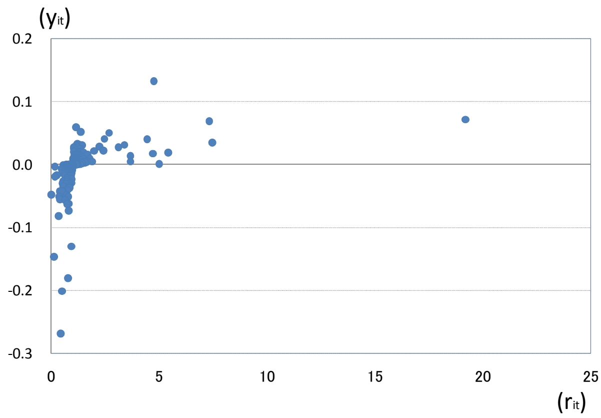
(1) Large enterprises of manufacturing



(2) Medium-sized enterprises of manufacturing



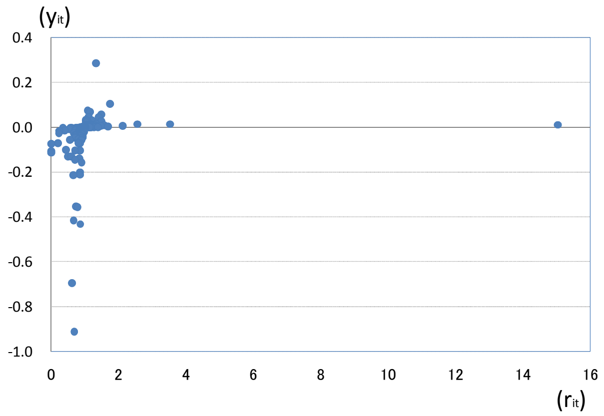
(3) Small enterprises of manufacturing



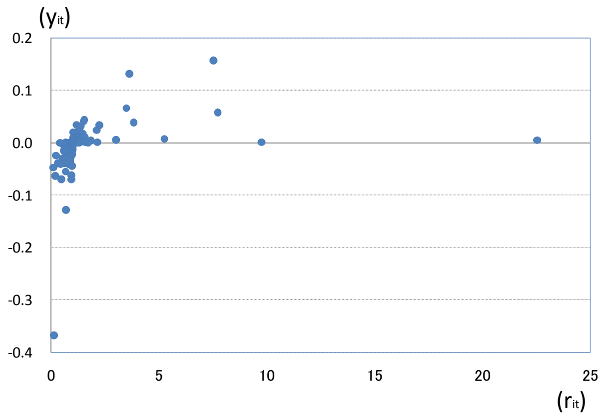
Note: Revision rate of Sales in 2009 from June survey, 2009

Figure 2-2: Relationship between r_i and y_{it}

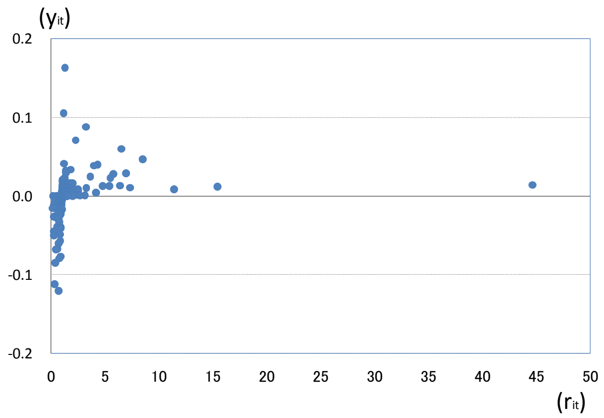
(1) Large enterprises of nonmanufacturing



(2) Medium-sized enterprises of nonmanufacturing



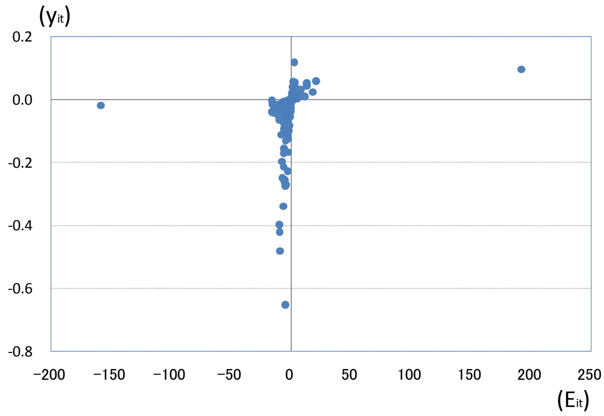
(3) Small enterprises of nonmanufacturing



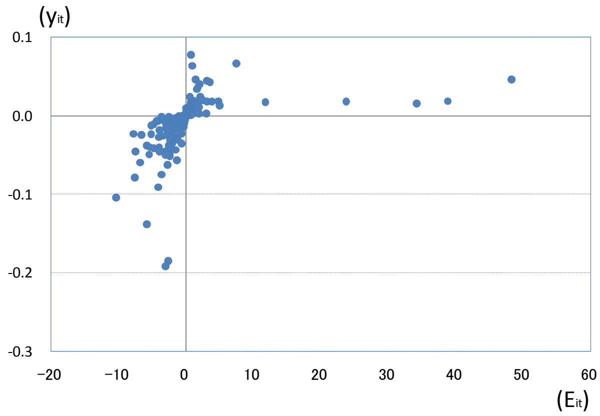
Note: Revision rate of Sales in 2009 from June survey, 2009

Figure 3-1: Relationship between E_{it} and y_{it}

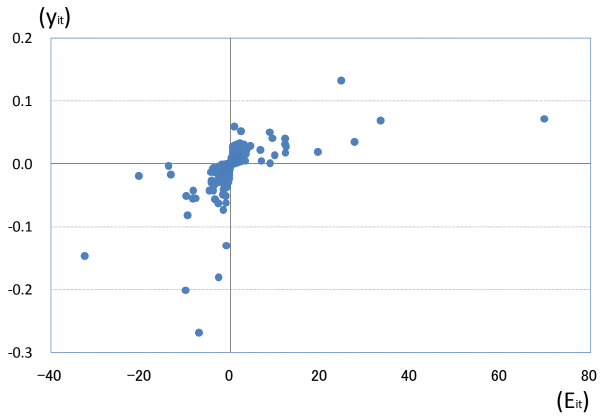
(1) Large enterprises of manufacturing



(2) Medium-sized enterprises of manufacturing



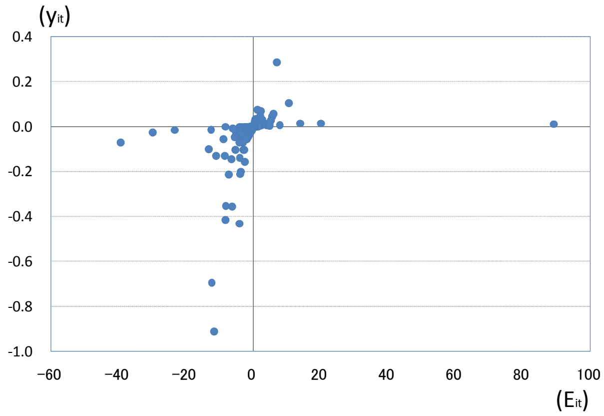
(3) Small enterprises of manufacturing



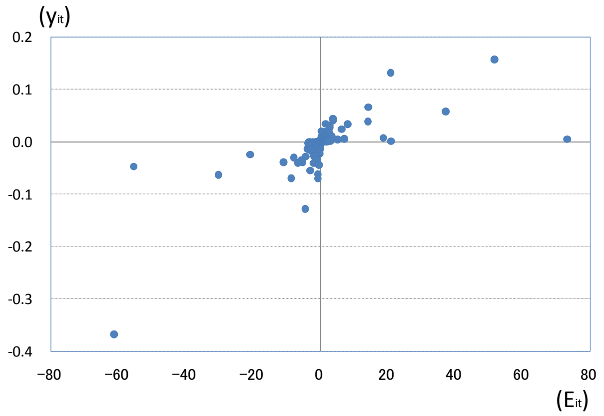
Note: Revision rate of Sales in 2009 from June survey, 2009

Figure 3-2: Relationship between E_{it} and y_{it}

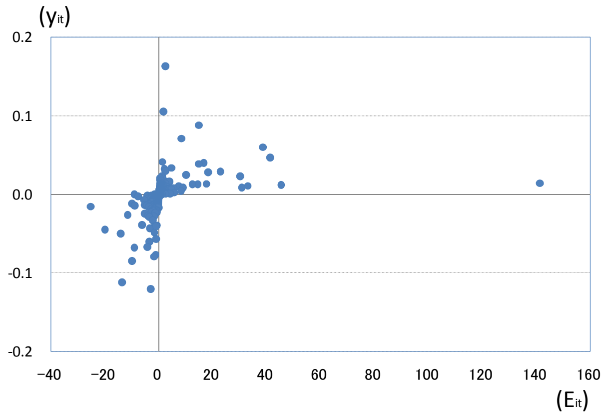
(1) Large enterprises of nonmanufacturing



(2) Medium-sized enterprises of nonmanufacturing[3]



(3) Small enterprises of nonmanufacturing



Note: Revision rate of Sales in 2009 from June survey, 2009

---

[3] One data, E_{it}: 26535, y_{it}: 0.12, is excluded.

Figure 4: Method to detect outlier

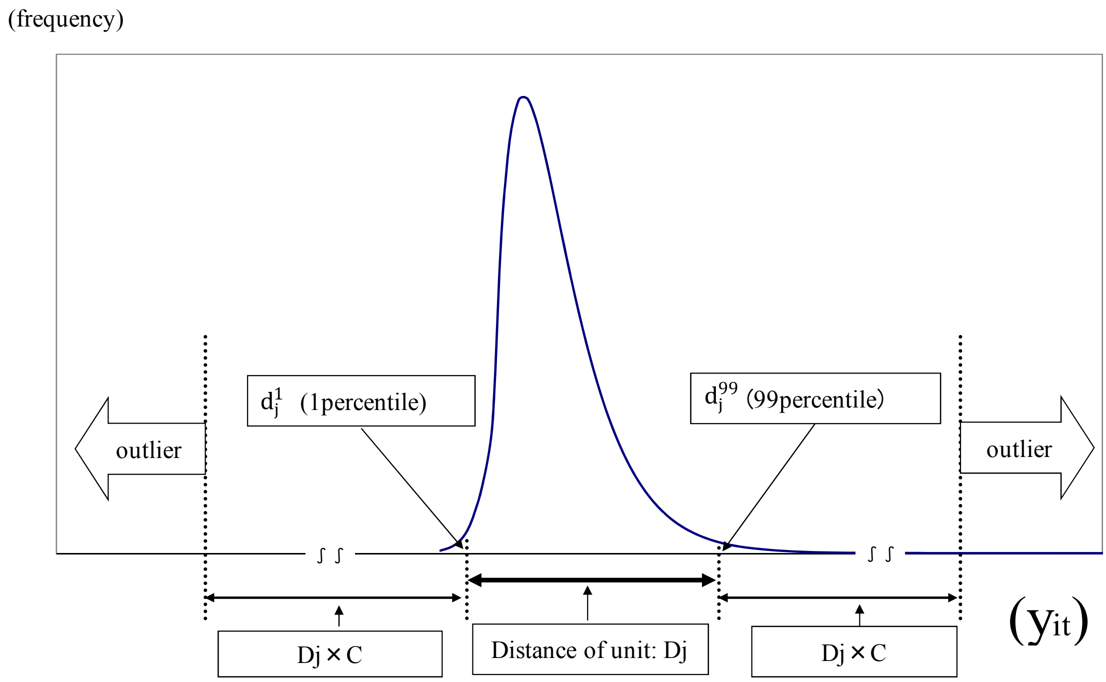
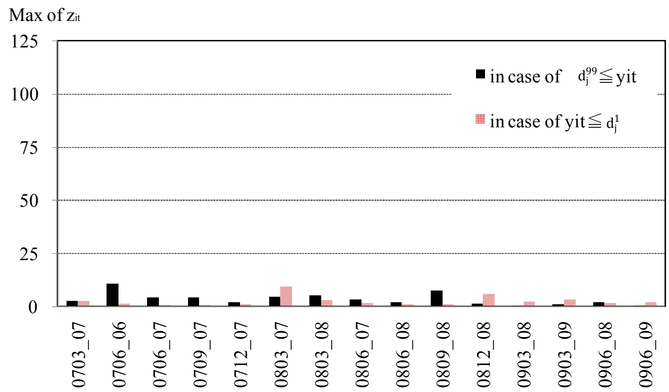
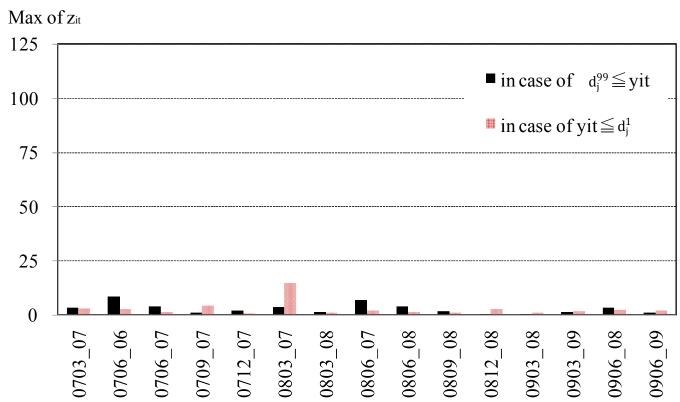


Figure 5-1: Max of z_{it}, indicator of outlier detection

(1) Sales: Large enterprises of manufacturing



(2) Sales: Medium-sized enterprises of manufacturing



(3) Sales: Small enterprises of manufacturing

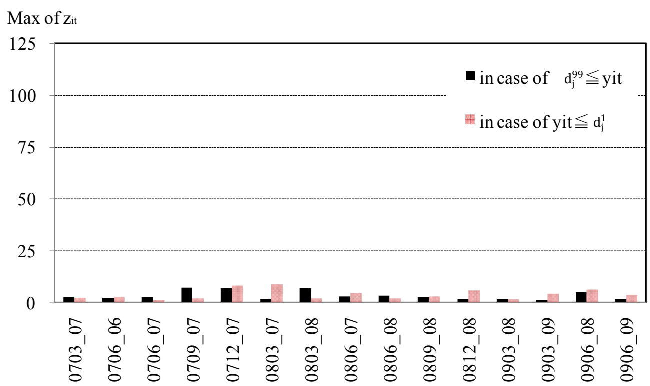
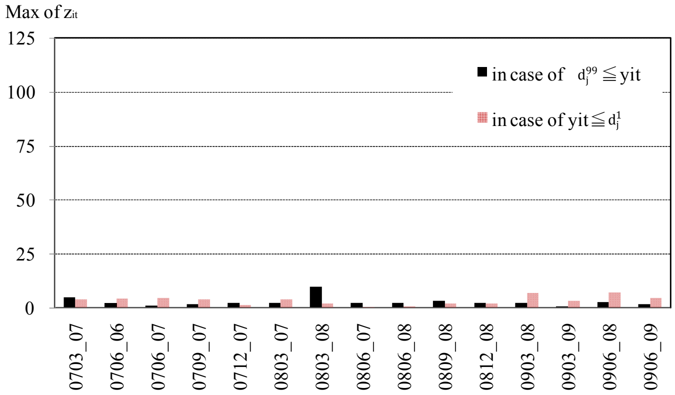
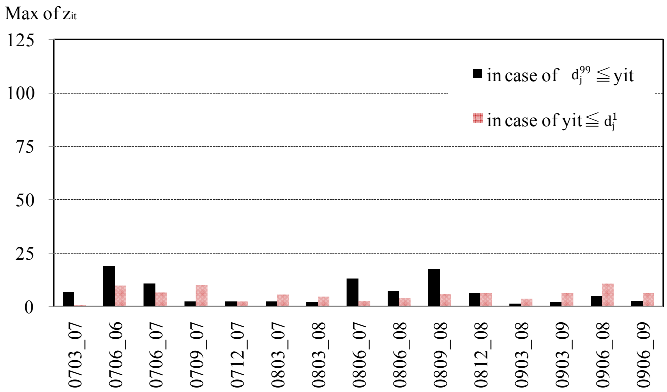


Figure 5-2: Max of z_{it}, indicator of outlier detection

(1) Sales: Large enterprises of nonmanufacturing



(2) Sales: Medium-sized enterprises of nonmanufacturing



(3) Sales: Small enterprises of nonmanufacturing

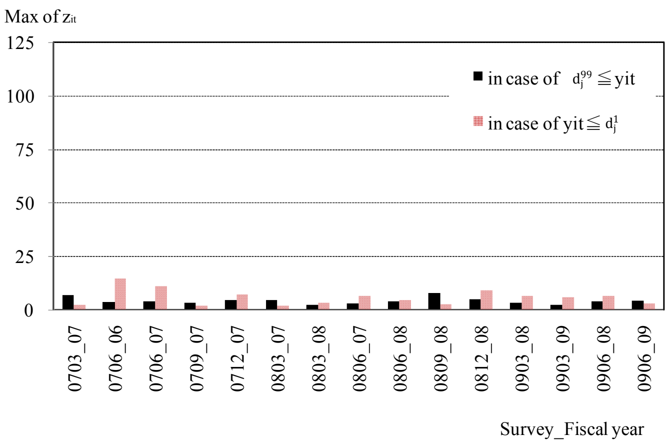
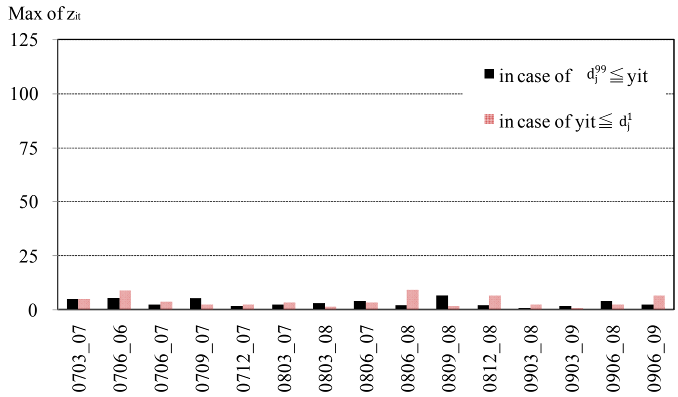
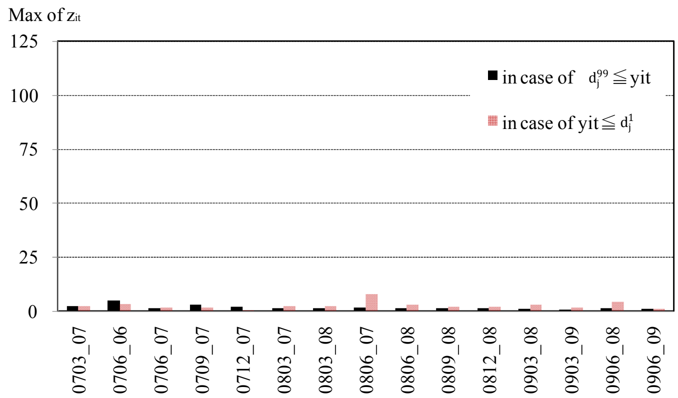


Figure 5-3: Max of z_{it}, indicator of outlier detection

(1) Current profits: Large enterprises of manufacturing



(2) Current profits: Medium-sized enterprises of manufacturing



(3) Current profits: Small enterprises of manufacturing

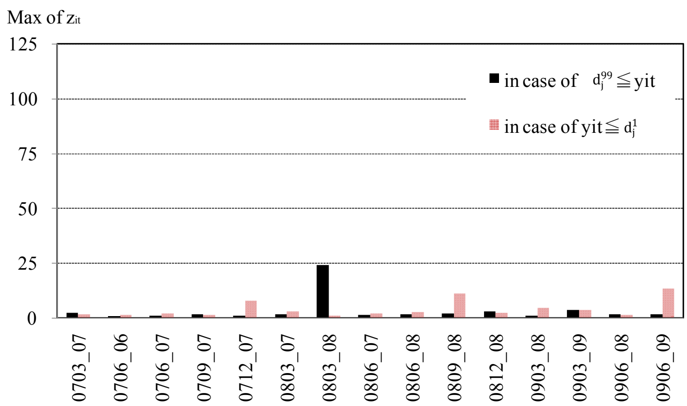
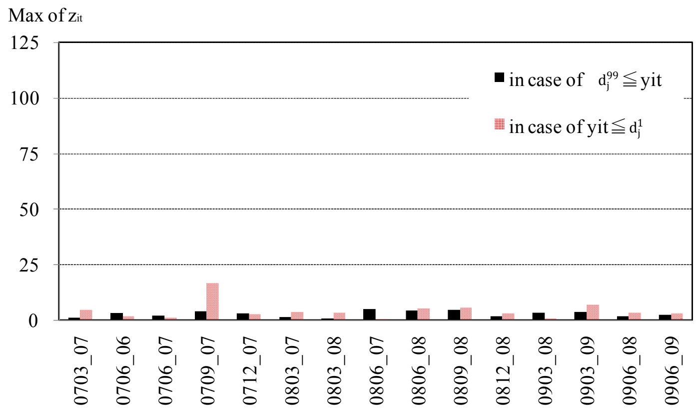
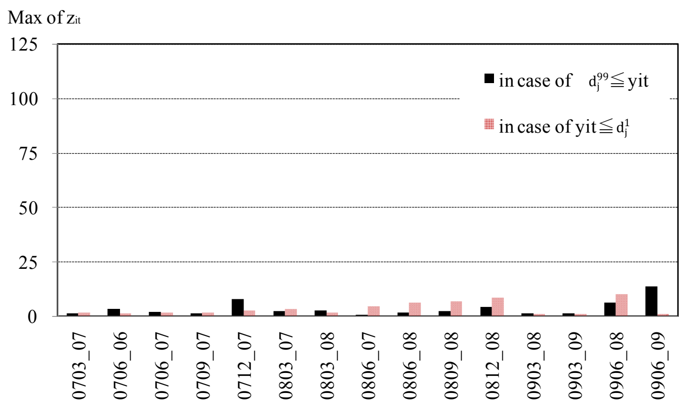


Figure 5-4: Max of z_{it}, indicator of outlier detection

(1) Current profits: Large enterprises of nonmanufacturing



(2) Current profits: Medium-sized enterprises of nonmanufacturing



(3) Current profits: Small enterprises of nonmanufacturing

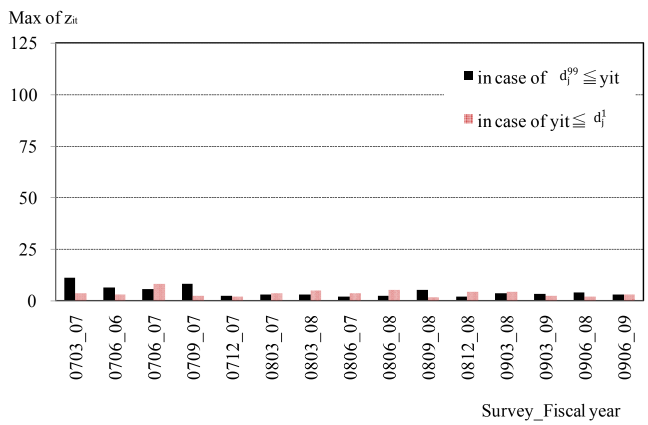
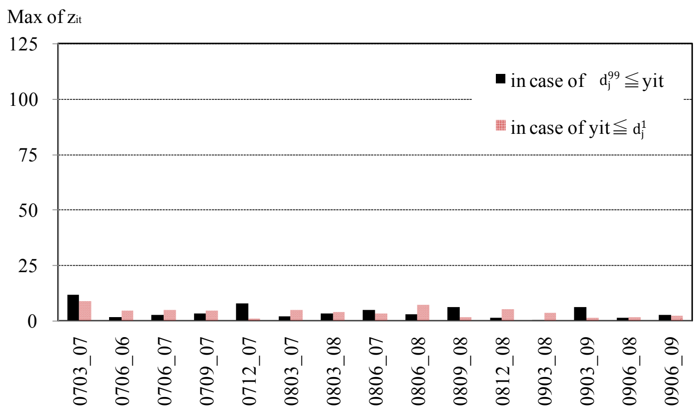
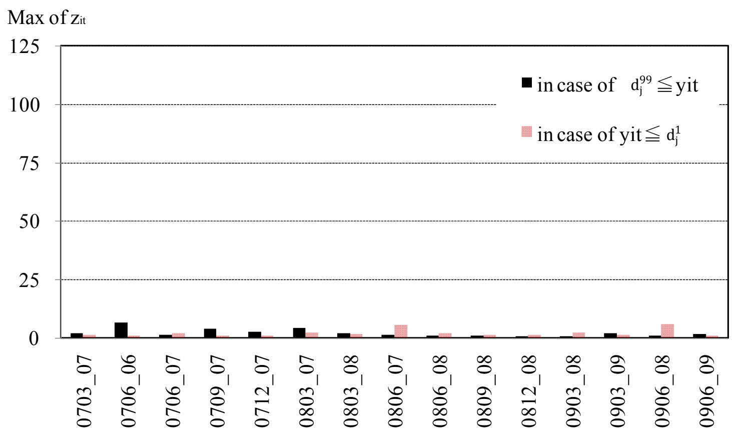


Figure 5-5: Max of z_{it}, indicator of outlier detection

(1) Net income: Large enterprises of manufacturing



(2) Net income: Medium-sized enterprises of manufacturing



(3) Net income: Small enterprises of manufacturing

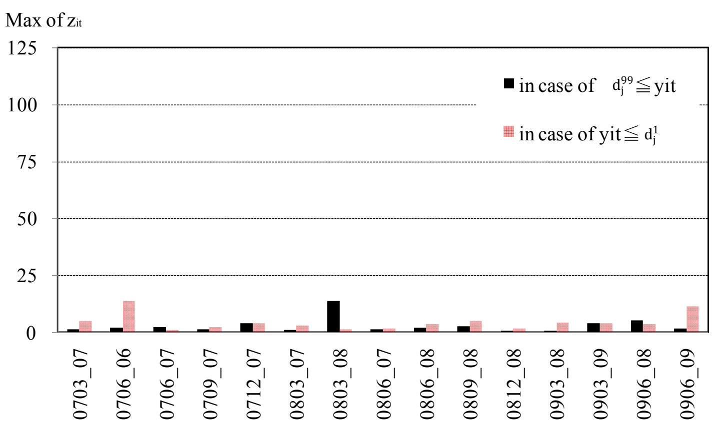
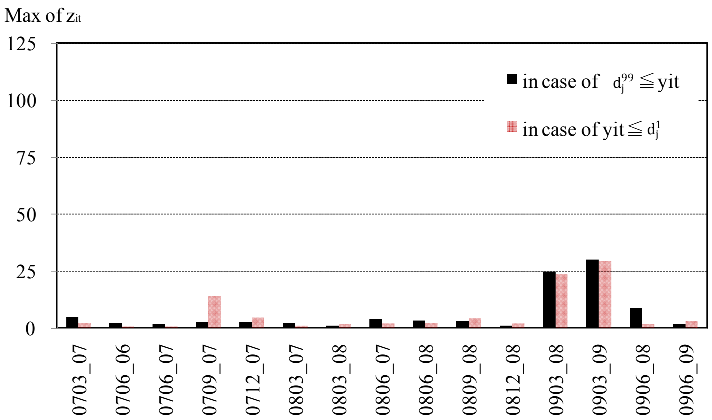
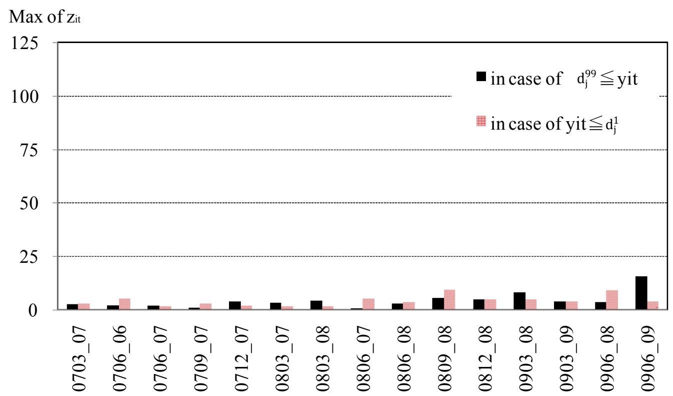


Figure 5-6: Max of z_{it}, indicator of outlier detection

(1) Net income: Large enterprises of nonmanufacturing



(2) Net income: Medium-sized enterprises of nonmanufacturing



(3) Net income: Small enterprises of nonmanufacturing

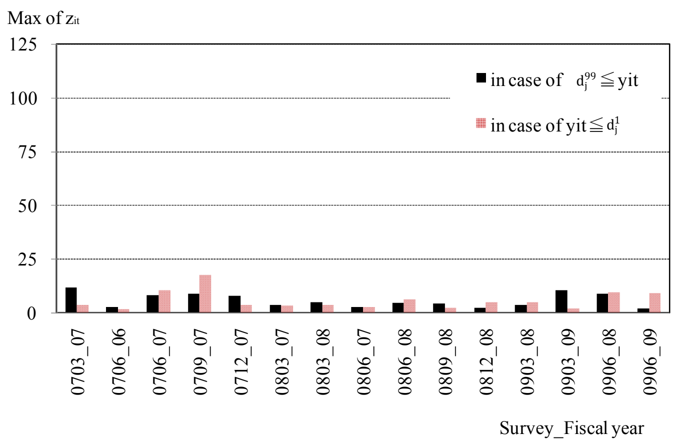
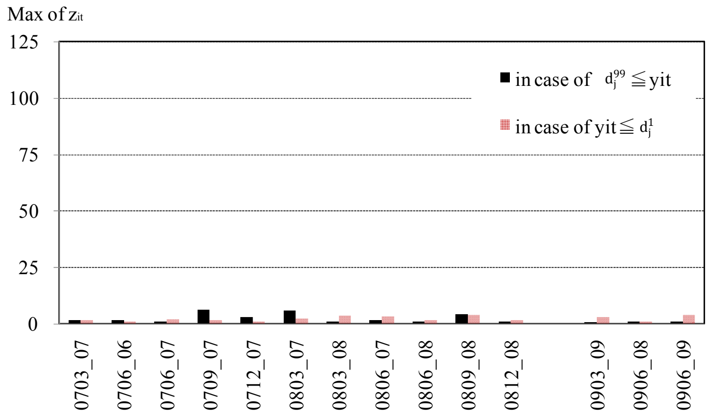
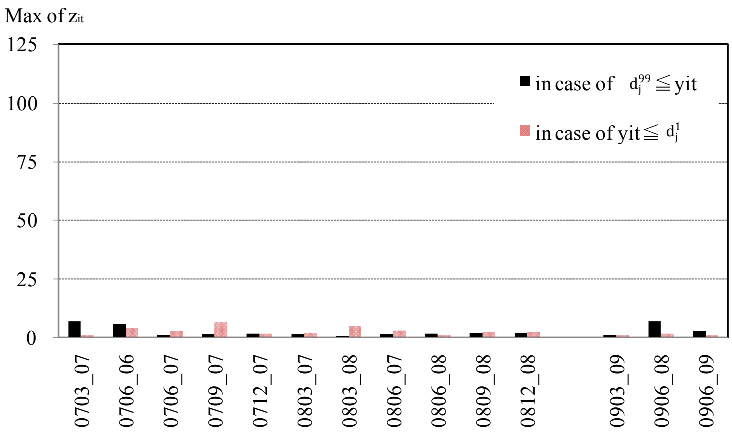


Figure 5-7: Max of z_{it}, indicator of outlier detection

(1) Fixed investment: Large enterprises of manufacturing



(2) Fixed investment: Medium-sized enterprises of manufacturing



(3) Fixed investment: Small enterprises of manufacturing

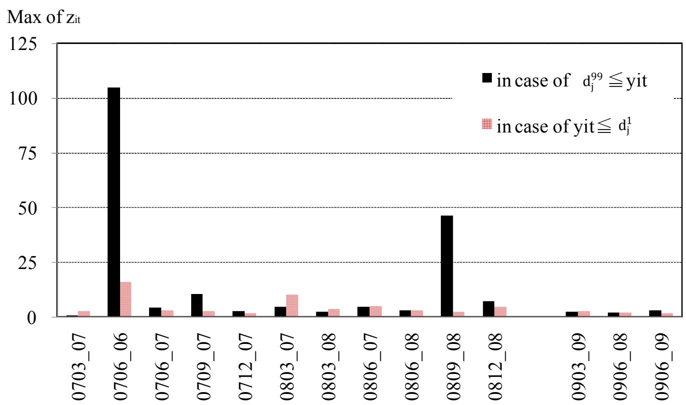
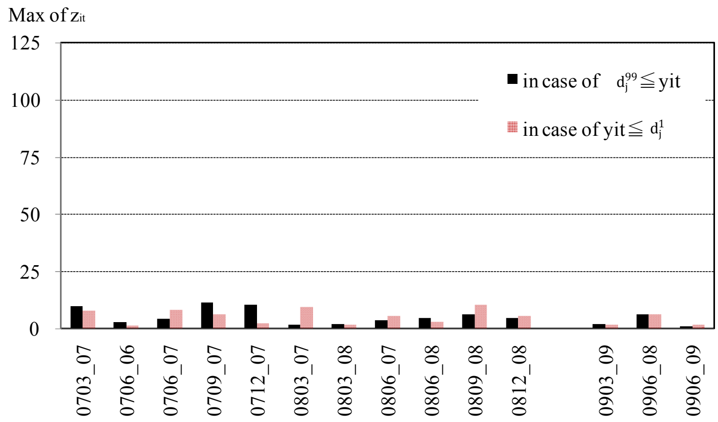
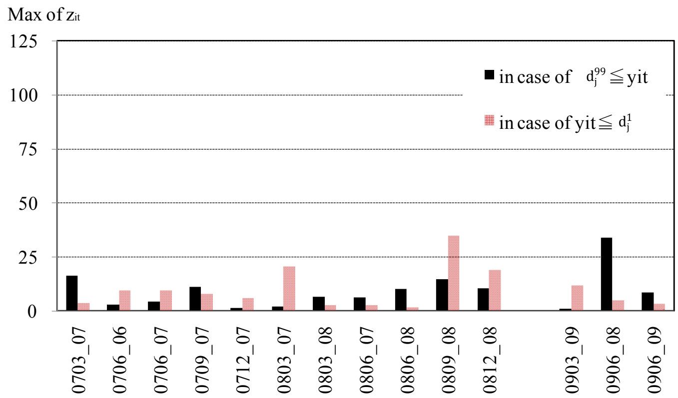


Figure 5-8: Max of z_{it}, indicator of outlier detection

(1) Fixed investment: Large enterprises of nonmanufacturing



(2) Fixed investment: Medium-sized enterprises of nonmanufacturing



(3) Fixed investment: Small enterprises of nonmanufacturing

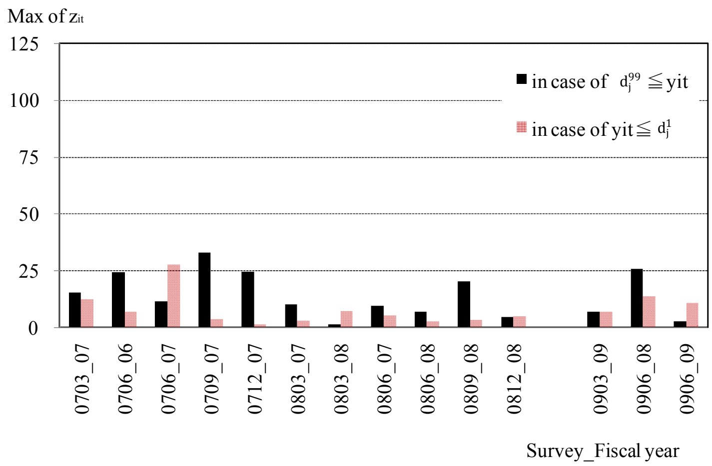
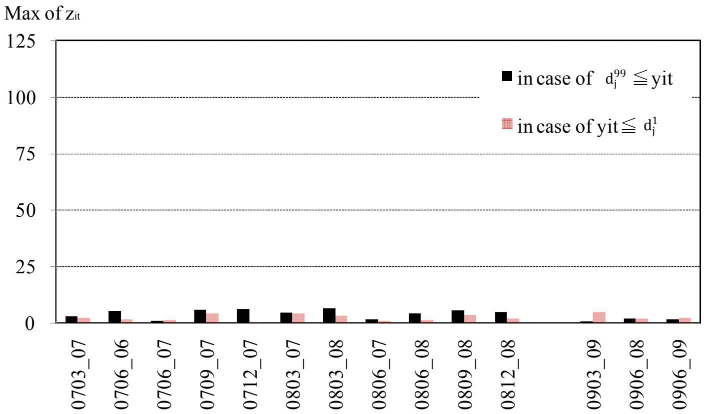
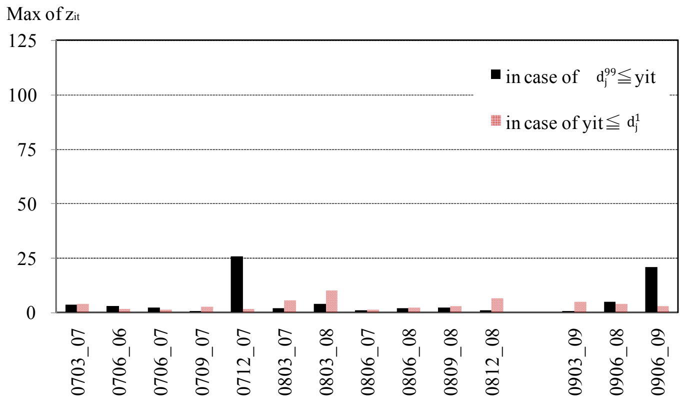


Figure 5-9: Max of z_{it}, indicator of outlier detection

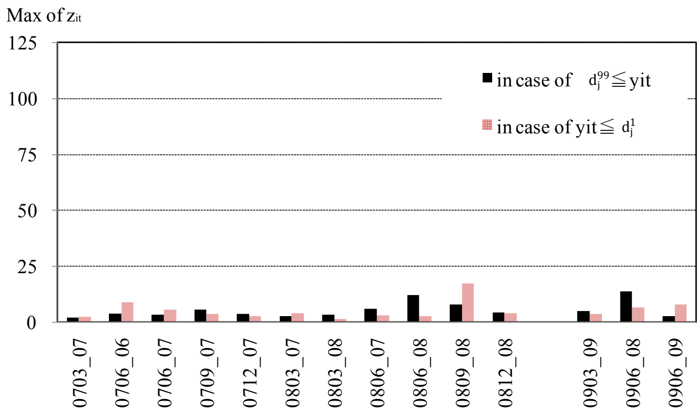
(1) Software investment: Large enterprises of manufacturing



(2) Software investment: Medium-sized enterprises of manufacturing

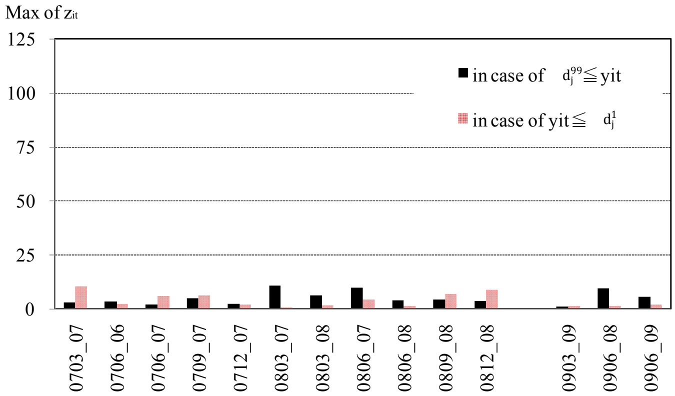


(3) Software investment: Small enterprises of manufacturing

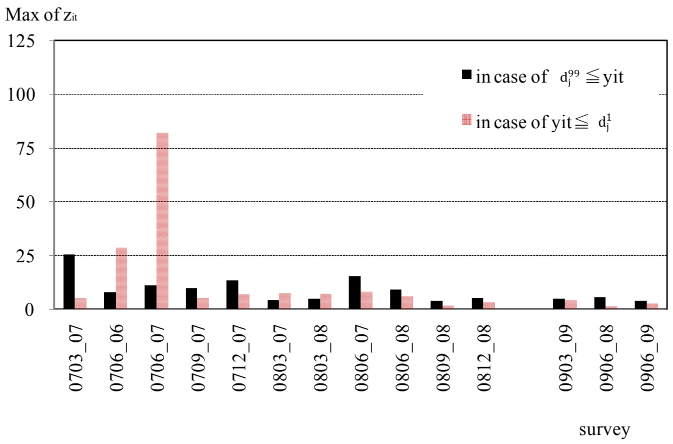


<u>Figure 5-10: Max of z_{it}, indicator of outlier detection</u>

(1) Software investment: Large enterprises of nonmanufacturing



(2) Software investment: Medium-sized enterprises of nonmanufacturing



(3) Software investment: Small enterprises of nonmanufacturing