Bank of Japan Working Paper Series

A Study of Missing Value Imputation in Business Surveys : The Case of the Short-Term Economic Survey of Enterprises in Japan (Tankan)

Takahiro Hirakawa*
takahiro.hirakawa@boj.or.jp

Junichiro Hatogai**
junichirou.hatogai@boj.or.jp

No.13-E-9
May 2013

Bank of Japan
2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-0021, Japan

* Research and Statistics Department

** Research and Statistics Department

A Study of Missing Value Imputation in Business Surveys : The Case of the Short-Term Economic Survey of Enterprises in Japan (*Tankan*)

Takahiro Hirakawa* and Junichiro Hatogai**

May 2013

[Abstract]

In business surveys, how to treat the data on which no responses are obtained (missing values) is an important theme in maintaining and improving the accuracy and reliability of statistics. The Short-Term Economic Survey of Enterprises in Japan (*Tankan*) enjoys a high response rate thanks to the cooperation of the enterprises it covers. Nevertheless, each survey inevitably includes items with missing values. When there are incidences of missing values in quantitative data items, such as sales and fixed investment, the current practice is to compile the survey result by imputing the missing value with the "previous fiscal year's value obtained from the non-responding enterprise."

This imputation method is thought to be appropriate when the economic environment is stable, as the change in the values between the previous term and the current term is relatively small. On the other hand, during a phase of sharp changes in the economic environment, the result of the compilation may diverge somewhat from the real perception of business conditions, as the previous term's value, which does not appropriately reflect the change during the period, is used without modification. Given the fact that in recent years, there has been a dramatic economic change, such as the one brought about by the Lehman shock, it is worth examining if there are imputation methods which produce a higher degree of accuracy.

Against this backdrop, this paper has compared the degrees of statistical accuracy of the current method of missing value imputation of the *Tankan* and its alternative methods, conducting simulations based on the data for the period from 2004 to 2010. As a result, it was found that for all of the

major survey items (i.e., fixed investment, sales and current profits), "there are alternative methods whose degree of accuracy is higher than or about the same as the imputation method now in use."

*Research and Statistics Department, Bank of Japan
(E-mail: takahiro. hirakawa @ boj.or.jp)

**Research and Statistics Department, Bank of Japan
(E-mail: junichirou. hatogai @ boj.or.jp)

This paper was written based on the report made by the authors at the Fifth Joint EC-OECD Workshop on Recent Developments in Business and Consumer Surveys jointly sponsored by the European Commission (EC) and the Organization for Economic Cooperation and Development (OECD) held in Brussels on November 17-18, 2011.

The authors would like to acknowledge and express gratitude for the valuable comments received from Messrs. Kosuke Aoki (the University of Tokyo), Kiyohito Utsunomiya (Kansai University) and Hiroshi Saigo (Waseda University) as well as Messrs. Eiji Maeda, Chihiro Sakuraba, Toshitaka Sekine, Koichiro Kamada, Shuji Kobayakawa, Seisaku Kameda and others of the Bank of Japan in preparing this paper. They would also like to thank the staff members of the Economic Statistics Division, Research and Statistics Department, with whom they have discussed in depth the missing value imputation methods and simulations examined in this paper.

1. Introduction

The Bank of Japan conducts the Short-Term Economic Survey of Enterprises in Japan (hereafter to be referred to as the "*Tankan*"), which is expected to provide an accurate picture of business conditions of enterprises in Japan, and thus to contribute to the appropriate implementation of monetary policy. The *Tankan* survey sends questionnaires to approximately 11,000 enterprises chosen as samples from among private enterprises in Japan and obtains responses on items such as sentiment on business conditions (judgement survey) and items on financial quantitative data, such as sales and profit plans (quantitative data items).

Thanks to the cooperation of the enterprises under survey, the *Tankan* has so far achieved high response rates. Nevertheless, in some surveys, non-negligible numbers of respondents send no responses to some of the surveyed items. In statistical surveys, the data which have not been obtained are called "missing values," and the insertion of values which are thought to be close to the missing values in their place is called "missing value imputation." At present, the *Tankan* imputes "the latest available value[1] reported by the enterprise which failed to respond in the current survey (non-respondent)" when a missing value is observed for a quantitative data item, while it does not make imputation for judgement survey.

The current method is deemed appropriate when the economic conditions are stable, as the change in the value between the previous term and the current term is expected to be small in these cases. On the other hand, when the economic conditions change drastically, the aggregated results obtained based on the current method could diverge from the actual perception of business conditions, because the current method uses the value reported in the previous term, which does not appropriately reflect possible changes of the economic conditions from the previous to the current term. Given the drastic change in the economic environment in recent years, such

[1] The latest available value could be the previous year's value or the current year's value obtained from the previous survey depending on the timing of the survey and the data availability as explained in detail in 1.(3).

as the Lehman shock, it would be worth exploring an imputation method with a higher degree of accuracy than the current method. Against this backdrop, this paper examines alternatives for the missing value imputation method currently in use for the *Tankan*.

This paper is organized as follows. Section 2 provides an overview of the recent incidences of missing values in the *Tankan*, followed by a summary of representative missing value imputation methods and a brief discussion about the motivation to conduct this study. Section 3 examines two alternative methods of missing value imputation for major survey items of the *Tankan* and lays out options to apply these alternative methods to each survey item. Section 4 compares the degree of accuracy of imputation of alternative methods with that of the current method based on the simulation using the *Tankan* data. Section 5 presents a summary of the study.

2. The incidences of missing values in the *Tankan* survey and the question this study attempts to answer

In order to grasp enterprises' behavior from various angles, the *Tankan* sets out a large number of survey items. They comprise items on business sentiment, such as on business conditions (henceforth "judgement survey"), and quantitative financial data, such as sales and profit plans (henceforth "quantitative data items") (Exhibit 1-1). The judgement survey asks enterprises' views on business conditions, as well as supply and demand conditions for products and services, inventory levels, production capacity and employment conditions[2]. The quantitative data survey queries actual values and projections for the fiscal year on sales, profits, fixed investment and others, and actual values on quarterly items, such as assets and liabilities.

Thanks to the cooperation of the enterprises under survey, the *Tankan* has

[2] The judgement survey releases enterprises' views in the form of a diffusion index. For example, the judgement of business conditions is expressed by the diffusion index obtained by subtracting the percentage share of enterprises responding "unfavorable" from that of enterprises responding "favorable."

so far enjoyed high response rates on an array of these questions. Nevertheless, in some surveys no responses are obtained on projections for the new fiscal year.

This subsection first explains how frequently the enterprises send no responses to the *Tankan* surveys. The *Tankan* maintains high response rates compared with other statistical surveys. However, in March surveys, which ask projections for the new fiscal year, the missing value rates tend to be relatively high. Then a brief survey of representative missing value imputation methods will be provided. Lastly, we will discuss how the imputation methods used for the *Tankan* have evolved, as well as the challenges this study attempts to address.

(1) The missing values in the *Tankan* survey

In the *Tankan* survey, the response rates for the judgement surveys are quite high. This is partly because respondents have been asked to cooperate on a continuing basis. On the other hand, missing values are not rare for quantitative data items, especially in March surveys, which ask projections for the new fiscal year for the first time. This is because enterprises, large ones in particular, tend to refrain from responding as they are in the process of making business plans.

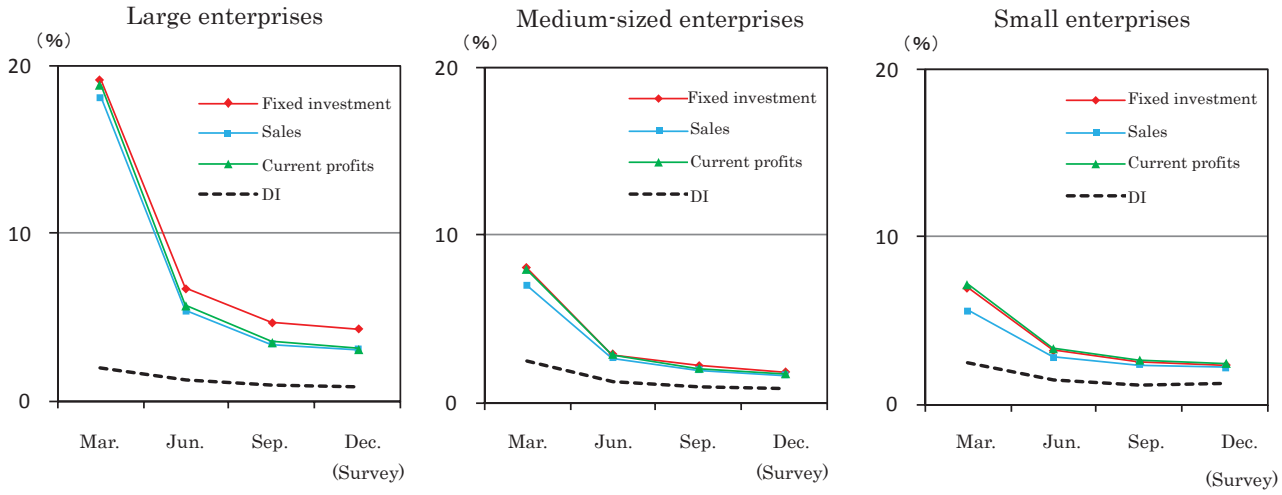
The charts below show the missing value rates (i.e., the proportion enterprises whose responses are other than valid responses among the total number of enterprises under survey) for annual projection items such as the amount of fixed investment, sales and current profits and business conditions judgement diffusion index by size of enterprises[3]. The missing value rates for business conditions judgement diffusion index are extremely low and shows no marked variance depending on the timing of the survey or on the size of responding enterprises[4]. On the other hand, the missing

[3] Fixed investment in this paper includes land purchasing expenses but exclude software investment. Figures for fiscal 2009 and later are under the new lease accounting standard.

[4] Generally, the released "response rate" is the effective response rate of business conditions diffusion index computed as the number of enterprise which gave effective responses divided by the number of enterprises under survey, which equals one minus

value rates of annual projection items are relatively high at about 15-20 percent for large enterprises and about 5-10 percent for medium-sized and small enterprises in March survey, though they fall to about 3-5 percent in June and later surveys.

[Missing value rate]



(Note 1) Missing value rate = 1 - (Number of enterprises giving effective response/Number of enterprises under survey)
(Note 2) The average for 2009 - 2011
(Note 3) DI : Business conditions DI

(2) The representative missing value imputation methods

How, then, are these missing values generally treated in the compilation of statistics? The following is an overview of missing value imputation methods, which are a major theme of study in the execution of a statistical survey.

In the area of statistical surveys, various imputation methods are suggested depending on data characteristics or sampling methods [5]. Here, representative methods being widely employed in practice are classified

---

missing value rates.

[5] The following is a summary based on Utsunomiya and Sonoda (2001), Ton et al (2010), and others. As stated in Utsunomiya and Sonoda (2001), there is no clear consensus on the classification of missing value imputation methods. Here, the authors have classified primarily those methods which are being used in practice, such as for the *Tankan* and other statistics in Japan and abroad.

based on whether the data used for the imputation are limited to those for the survey item for which the value is missing (henceforth "missing value item") or the data used for the imputation include those for survey items other than the missing value item. Then, the methods belonging to the latter are further classified into (a) methods which use the data obtained from responding enterprises and (b) methods using the data obtained from the non-respondent (henceforth "missing value enterprise.")

According to this classification of missing value imputation methods, the methods which use the data obtained from responding enterprises—summarized as (a) in the table below—include "Random Hot Deck Donor Imputation" ([i] in the table below) and "mean imputation" ([ii] in the table below). The former uses the value obtained from a responding enterprise chosen at random to impute the missing value, while the latter uses the average value for a group of responding enterprises whose attributes are similar to those of the non-responding enterprise.

The methods which use the data obtained from the non-respondent—summarized as (b) in the table below— include "previous value imputation" ([iii] in the table below) and "growth rate imputation" ([iv] in the table below). The former is currently in use for the *Tankan*. It uses the latest available value obtained from the non-respondent to impute the missing value, while the latter uses the latest available value of the non-respondent taking account of the rate of growth from the previous term obtained from the responding enterprises (growth rate) to impute the missing value for the current term.

The imputation method which uses the data other than those for the missing item includes "ratio imputation[6]" ([v] in the table below). Under the "ratio imputation" method, the ratio of the value for the missing item to the known value for another item (the ratio between items) is used to multiply the value for the known item to obtain the missing value.

---

[6] The "ratio imputation" in a broader definition refers to any missing value imputation methods using ratio between items.

[Representative missing value imputation methods]

Data used for imputation		Missing value imputation methods	
		Example	Explanation
Data for missing item only	(a)Obtained from responding enterprises	[i] Random Hot Deck Donor Imputation	Directly uses the value obtained from a responding enterprise chosen at random
		[ii]Mean imputation	Uses the average for a group of enterprises with similar characteristics to the non-respondent
	(b)Obtained from the non-respondent	[iii]Previous value imputation (Now used for the *Tankan*)	Uses the previous term's value of the non-respondent without modification
		[iv]Growth rate imputation	Previous term's value of the non-respondent adjusted by "the growth rate from the previous term" obtained from responding enterprises
Data for missing item+ Data for other items		[v]Ratio imputation	The ratio of the missing item to another item is used to impute missing value

Compared in this paper

(3) Changes in the missing value imputation methods used for the *Tankan* and the motivation of this study

Based on the above summary of imputation methods, this paper will now review the developments in the missing value imputation methods used for the *Tankan*. Until 2004, the *Tankan* had used the "mean imputation." Based on the results of the study in Utsunomiya and Sonoda (2001), however, the "mean imputation" has been replaced by the "previous value imputation" since 2004.

With a view to conducting a thorough examination on the accuracy of the methods using exclusively the data for the missing item, Utsunomiya and Sonoda (2001) compared the degrees of statistical accuracy of the previous value imputation, growth rate imputation and the mean imputation, the last being the one in use for the *Tankan* at the time of the study. The Random Hot Deck Donor Imputation was excluded from the study as the large variance of the *Tankan* data could make it difficult for a value obtained from an enterprise chosen at random to represent the attributes of the non-respondent. Based on a number of simulations, the paper found that the statistical accuracy is higher for the previous value imputation and the growth rate imputation, both of which benefit from the characteristics of

time-series data, than for the mean imputation, which is susceptible to the variance of the data. Between the previous value imputation and the growth rate imputation, Utsunomiya and Sonoda (2001) concluded that, for the period from fiscal 1999 to fiscal 2000 which the paper focused on, overall the previous value imputation showed a better performance. On the basis of these results, the *Tankan* has adopted the previous value imputation rather than the mean imputation since 2004.

At present, the previous value imputation method is used only for annual projections among the quantitative data items[7]. Specifically, this method works as follows. When data for the new fiscal year are not obtained in the March survey, the value reported by the non-respondent in the previous fiscal year is used to impute the missing value. When data for the fiscal year are not obtained in the surveys in June or later, the relevant enterprise's data for the current fiscal year used in the compilation of the previous survey are used.

Thus, the previous value method now in use uses the latest available value reported by the non-respondent to impute the missing value. When the economic conditions are stable, this method is deemed appropriate, as it can be assumed that the data from the same enterprise do not change greatly between the previous and current survey in such a period. However, when an event hits an economy which could drastically change the economic conditions, such as the Lehman shock, this method may not accurately reflect the projection for the population enterprises, because the data for the previous survey are used without modification[8].

Against this backdrop, this study has examined whether or not there are alternatives to the previous value imputation in light of standard statistical criteria. The following study focuses on the three major items which attract the most attention from the users of the statistics (i.e., fixed investment,

---

[7] Among the quantitative data items, the missing value imputation is not made for quarterly data or for the number of new graduates hired.

[8] Utsunomiya and Sonoda (2001), which preferred the previous value imputation, states, "When the economic phase is clear,. . . the 'growth rate imputation' may further increase the degree of accuracy."

sales and current profits)[9]. The Random Hot Deck Donor Imputation and the mean imputation, which had been judged inappropriate in past studies, were excluded also from this study. Thus, the growth rate imputation and the ratio imputation were compared with the previous value imputation now in use, in terms of statistical accuracy[10].

3. An examination of alternative methods

This section will examine the alternative methods by comparing them with the previous value imputation.

(1) An examination of alternative methods for the three major survey items

The previous value imputation can be expressed by the following formula. Assuming that no response was obtained for the period *t*, from the enterprise *i* on survey item *y*, the missing value is to be expressed as $y_i^{(t)}$. And if the missing value to be imputed (henceforth "imputed value") is to be expressed as $\tilde{y}_i^{(t)}$, the previous value imputation can be expressed as follows.

$$\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{y_i^{(t-1)}}_{\text{The latest available value of the non-respondent}}$$

This approach under the previous value imputation assumes that the latest available value of the non-respondent does not diverge greatly from the missing value for the current survey and thus is a good proxy for the missing value. However, this assumption does not necessarily hold good when the economic conditions change drastically in a short period of time.

---

[9] According to a questionnaire conducted in 2004 by the Japan Federation of Economic Organizations, for example, most frequently used data out of the quantitative data of the *Tankan* were "fixed investment projections, etc., " followed by "sales and profit projections" (Exhibit 1-2).

[10] In addition to the "representative missing value imputation methods" shown under the previous item, there is the regression imputation, which uses the estimated value obtained from a regression analysis by assuming a model which explains the missing value. However, it was excluded from this study, because the cost of building and maintaining the model and calculating the results for each survey is prohibitingly high.

(The first alternative method)

The first alternative, the growth rate imputation, is expressed in a formula below. As shown in the formula, this method adds an element which helps capture the change in the data of the non-respondent, i.e., an approximate value of the rate of growth of the non-respondent, to the previous value imputation (Formula 3-1). Specifically, the approximate value of the growth rate of the non-respondent can be obtained by establishing a group of responding enterprises whose characteristics are similar to those of the non-respondent and by using the rate of change in the sum of the values given by the enterprises in this group as the imputed value[11].

$$\tilde{y}_i^{(t)} = y_i^{(t-1)} \times \underbrace{\frac{\sum y_k^{(t)}}{\sum y_k^{(t-1)}}}_{\text{Approximate value of the growth rate of the non-respondent computed as the rate of change of the sum of the values reported by the "group of responding enterprises whose characteristics are similar to those of the non-respondent"}} \quad \begin{array}{l} \text{...Value of the responding enterprises } (k \neq i) \text{ (current survey)} \\ \text{...The latest available value of the responding enterprises } (k \neq i) \end{array} \qquad (3\text{-}1)$$

By defining the "growth rate imputation" as the method under which the imputed value is the latest available value reported by the non-respondent multiplied by the approximate value of the growth rate of the non-respondent, it can capture to some extent the change from the previous survey to this survey, which was not reflected under the previous value imputation method.

---

[11] Another approach is to use the "average of the rates of change of individual enterprises in the group of responding enterprises whose characteristics are similar to those of the non-respondent" (See the formula below. n denotes the number of enterprises in the group. It is possible to adopt, in addition to the "mean value," the "median value" or the "mode value" of the rates of change of individual enterprises).

$$\tilde{y}_i^{(t)} = y_i^{(t-1)} \times \frac{1}{n}\sum \frac{y_k^{(t)}}{y_k^{(t-1)}}$$

However, given the present *Tankan* system, the burden of developing the system for the introduction of this method is excessive. Therefore, this study will adopt the Formula 3-1.

(The second alternative method)

Out of the three major survey items this study discusses, current profits can be negative, in other words, the respondent enterprises could suffer losses. Therefore, if the growth rate imputation is applied to current profits when the average amount of profits of the responding enterprises changes from a positive number (i.e., generating profits) to a negative number (i.e., suffering losses), or vice versa, the imputed value can greatly diverge from the reality[12]. Therefore, to use the growth rate imputation for current profits would not be appropriate.

Hence, as the second alternative imputation method for current profits, this paper will apply the "inter-item ratio imputation," which utilizes the relationship between the missing item and other survey items. Among several possible approaches of the inter-item ration imputation, this paper employs a method to obtain the imputed value by multiplying the current survey's value of sales obtained from the non-respondent[13] by the latest available current profits to sales ratio of the non-respondent. This can be expressed by the following formula (x denotes the value other than for the missing item).

$$\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value of the current profits}} = \underbrace{x_i^{(t)}}_{\text{The current survey's value of sales of the non-respondent}} \times \underbrace{\frac{y_i^{(t-1)}}{x_i^{(t-1)}}}_{\text{The latest available value for the current profits to sales ratio}} \quad \begin{array}{l} \text{…The latest available value of the missing item of the non-respondent} \\ \text{…The latest available value of an item other than the missing item of the non-respondent} \end{array} \tag{3-2}$$

[12] For example, if the average current profits of the responding enterprises shifted from minus 1 in the previous term to plus 10 in the current term, the "growth rate" would be minus 10-fold. If the previous term's current profits of the non-responding enterprise were plus 3, under the growth rate imputation, the imputed value for the missing value (the current term's number for the non-responding enterprise) would be minus 30. Under this condition, while the average current profits of the responding enterprises whose characteristics are similar to those of the non-responding enterprise rose (from minus 1 to plus 10 [the margin of change of plus 11]), for the current term's current profits of the non-responding enterprise, the imputed value would be greatly lower than the previous term's value (from plus 3 to minus 30 [the margin of change of minus 33]).

[13] If both current profits and sales for this term show missing values, the value obtained by growth rate imputation [industry-size] (which will be discussed later) can be regarded as this term's values.

(2) Implementation of alternative methods

Based on the discussions in the subsection (1), the "growth rate imputation" is applied to fixed investment and sales and the "inter-item ratio imputation" is applied to current profits, and the results are compared with those of the "previous value imputation" in terms of the degree of accuracy (for a list of alternative methods, see Exhibits 2-1 and 2-2).

(The growth rate imputation applied to fixed investment and sales)

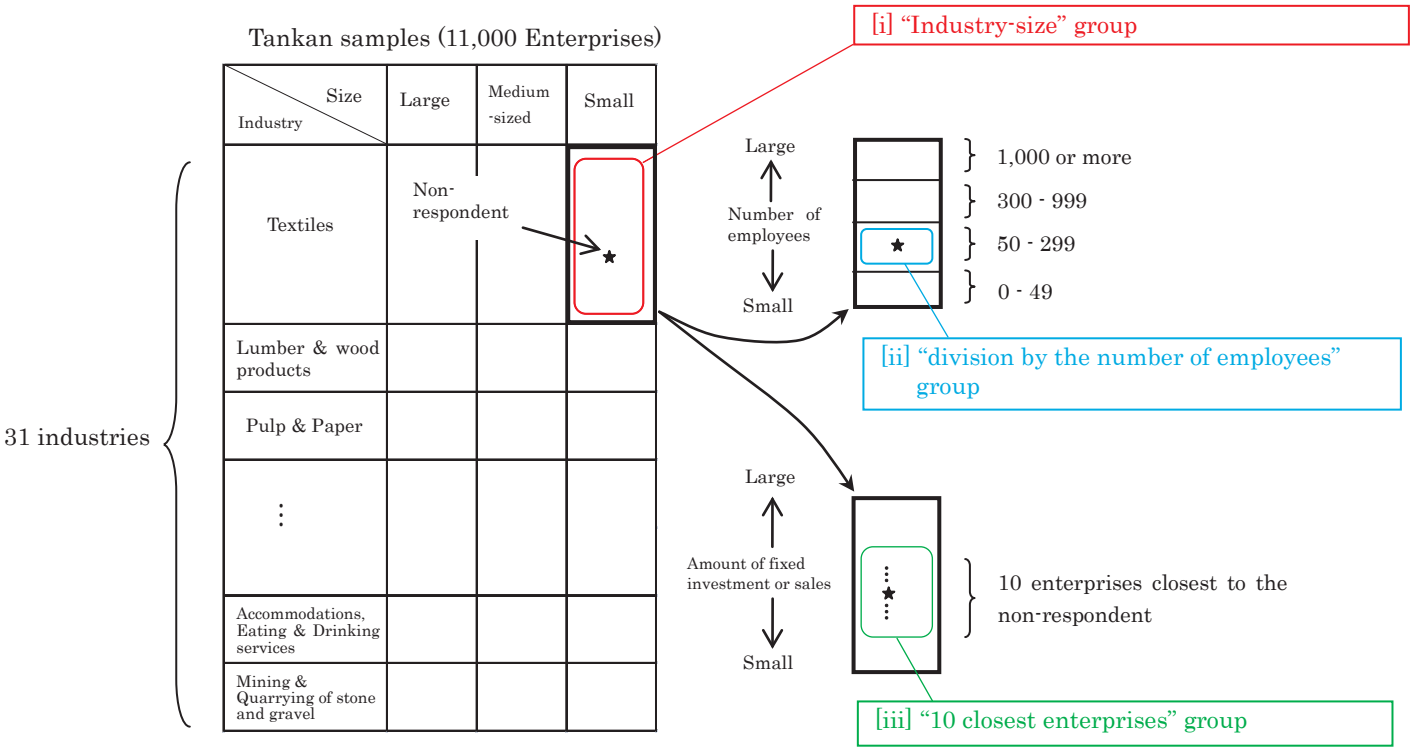
For fixed investment and sales, the growth rate imputation (Formula 3-1) is applied as an alternative to the previous value imputation.

In applying Formula 3-1, the degree of accuracy of the imputation hinges on how to delineate the group of responding enterprises whose characteristics are similar to those of the non-responding enterprise (henceforth "the group of comparable enterprises"). The narrower the scope of the group of comparable enterprises, the closer will be the characteristics of the chosen enterprises to those of the non-responding enterprise. On the other hand, as the number of enterprises in the group gets smaller, the impact of changes specific to certain enterprises may get larger. Hence, it is difficult to pre-determine the most appropriate scope of the group of comparable enterprises. Therefore, this study has examined using three groups of varying scopes.

As the *Tankan* data are compiled by the industry and size of enterprises, the good departing point is to formulate the group of comparable enterprises by collecting the enterprises in the same industry and size categories as the non-responding enterprise (henceforth "industry-size" group; [i] in the chart below). The scope of the "industry-size" group may be, however, too broad to meet the requirement to choose appropriately the enterprises which have the same quality as the non-responding enterprise. Therefore we have created two variations of the grouping method whose scopes are narrower than the simple "industry-size" group. One is the grouping method in which the "industry-size" group is divided on the basis of the number of employees

(henceforth “division by the number of employees” group; [ii] in the chart below)[14]. The other is the grouping method to choose the ten enterprises closest to the non-respondent, namely five enterprises above and five enterprises below the non-respondent on the list, in terms of the amount of fixed investment or sales for the previous fiscal year (henceforth “10 closest enterprises” group; [iii] in the chart below). Then the accuracy of statistical inference will be compared between the current method and each of these three alternative methods.

[The group of responding enterprises similar to the non-responding enterprise]



(Inter-item ratio imputation for current profits)

For current profits, the inter-item ratio imputation using the “current profits to sales” ratio (Formula 3-2) will be applied as an alternative to the previous value imputation.

[14] In principle, the division was made on the basis of the number of employees: “0 — 49,” “50 — 299,” “300 — 999” and “1,000 or more” to create four groups.

While this formula assumes that the ratio between survey items has not changed significantly between the previous and current survey, it is not self-evident that the change in the ratio between the previous and current survey was small. Therefore, in order to approximate as much as possible the latest available figure for the ratio between the survey items of the non-responding enterprise to the unknown ratio between the survey items of the non-responding enterprise for the current survey, two variations to Formula 3-2 were examined.

The first approach is to obtain data on the ratio between the survey items of enterprises other than the non-respondent from the same survey. To obtain the ratio between the survey items of the responding enterprises whose characteristics are similar to those of the non-respondent from the same survey, the group introduced in the examination of the growth rate imputation was used. In specific, this study has adopted the "average ratio between the survey items in the same survey of the "industry-size" group, to which the non-respondent belongs (Formula 3-3. Henceforth referred to as the "average response.")

$$\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{x_i^{(t)}}_{\text{The current survey's value for another item obtained from the non-respondent}} \times \underbrace{\frac{\sum y_k^{(t)}}{\sum x_k^{(t)}}}_{\text{The current survey's average ratio between the survey items of the "industry-size" group to which the non-respondent belongs}} \qquad (3\text{-}3)$$

Approximate value of "the current term's ratio between the survey items of the non-respondent"

The second approach is aimed at taking account of the change in the ratio between survey items from the previous survey to the current survey of the non-respondent. Specifically, this approach uses the latest available value of the ratio between the survey items of the non-respondent and the change (i.e., margin of increase) from the previous survey to the current survey in the average ratio between the items of the "industry-size" group to which the non-respondent belongs (Formula 3-4. Henceforth, "the margin of increase.")

$$\tilde{y}_i^{(t)} = x_i^{(t)} \times \left\{ \frac{y_i^{(t-1)}}{x_i^{(t-1)}} + \left(\frac{\sum y_k^{(t)}}{\sum x_k^{(t)}} - \frac{\sum y_k^{(t-1)}}{\sum x_k^{(t-1)}} \right) \right\} \quad (3\text{-}4)$$

Imputed value

The current survey's value for another survey item obtained from the non-respondent

The latest available value for the ratio between the survey items of the non-respondent

+

The margin of change (margin of increase) in the average ratio between the survey items between the previous term and the current term of the "industry-size" group to which the non-respondent belongs

The approximate value of the current survey's ratio between survey items of the non-respondent

4. A Comparison of the degrees of accuracy based on simulation

In this section, the degree of accuracy of the previous value imputation method now in use will be compared with that of the alternative methods discussed in Section 3 for the major three survey items of the *Tankan* (i.e., fixed investment, sales and current profits) based on the simulations using the data set of the *Tankan*.

An outline of the simulation method and the method of evaluation of its results is as follows. Simulation results are examined for each survey item to see whether the alternative methods improve the accuracy of imputation compared with the current method, followed by a summary of the examination. Lastly, the aggregate figures will be calculated based on the alternative imputation methods to be compared with those based on the current imputation method.

(1) An outline of the simulation method and the method of evaluation of simulation results

In order to compare the degrees of accuracy of imputation between the previous value imputation now in use and the alternative methods, Monte

Carlo simulation using the *Tankan* data set was conducted in the following manner.

[i] Based on the actual figures (the final data for the previous fiscal year obtained in June surveys) for fixed investment, sales and current profits for fiscal 2004 through fiscal 2010, a data set without missing values was created[15].

[ii] Then, the data set was divided into approximately 400 groups based on industry, size of capital and the number of employees[16].

[iii] Missing values were generated at random in each group based on the missing value rates for the projections for the following fiscal year in the March survey of each fiscal year[17], and then imputed values were obtained based on each imputation method[18].

[iv] After the missing values were imputed, the aggregate for the relevant item (i.e., the amount of population estimates) was obtained for each of the six groups of enterprises, i.e., large, medium-sized, and small enterprises for both manufacturing and non-manufacturing sectors.

[v] Steps [iii] and [iv] were repeated 500 times[19], and for the aggregate of the relevant item, the errors between the imputed values and the true values (i.e., the values in the original data set) were obtained as Relative Root Mean Squared Errors (RRMSEs).

The RRMSEs is one of the indicators of the error between the estimated value and the true value. It is the square root of the expected value of the

---

15 Specifically, from the data obtained in June surveys, when the actual figures for the previous year are finalized, a data set was created by excluding enterprises which did not respond or whose figures were not finalized.

16 These are the same as the "division by the number of employees" groups, which were created by dividing the "industry-size" group according to the number of employees.

17 For example, for the actual value for fiscal 2010, missing values were created at random based on the missing value rates in fiscal 2010 March survey. The missing value rates for March were used, because it is the highest in all of the surveys during the year. Hence, it was assumed that the simulation will demonstrate most clearly the difference among the imputation methods.

18 For each imputation method, imputation was made separately for the first half and the second half of the term.

19 An examination of Relative Root Mean Squared Errors (RRMSEs) for the cases in which the simulation was repeated 300 times, 400 times and 500 times has revealed that there was no significant difference among them. Therefore, it was concluded that results with sufficient convergence were obtained.

squared difference between the estimated value and the true value over the true value[20]. Therefore, the smaller this value, closer it is to the true value, and the accuracy of imputation is interpreted to be high. This study examined to what extent the alternative method has reduced the RRMSEs as compared with the current method by using the level of the RRMSEs as a yardstick to measure the statistical accuracy of the imputation method.

[An Outline of the Simulation]

Item	Fixed investment, sales and current profits
Data	Actual values for fiscal 2004 - fiscal 2010
Missing value rates	The missing value rates of the values for the new fiscal year in March surveys
No. of simulation	500 times
Evaluation method	The level of the RRMSEs for each of the six groups of enterprises comprising large, medium-sized, and small enterprises for both manufacturing and non-manufacturing sectors

(2) Simulation results and appraisal

(Fixed investment)

With respect to the amount of fixed investment, a comparison of the accuracy (Exhibit 3-1) of the abovementioned three growth rate imputation methods (Exhibit 2-1) found that the RRMSEs of the "division by the number of employees" was clearly larger than the other two in a number of cases.

---

[20] The process for obtaining the RRMSE is shown below. First, Root Mean Squared Error (RMSE) is defined as follows as the square root of the expected value of the "squared error between the estimated value and the true value ($\hat{\theta}$ denotes the estimated value, while θ denotes the true value)."

$$RMSE(\hat{\theta}) = \sqrt{E\{(\hat{\theta} - \theta)^2\}}$$

RRMSE, obtained by dividing this RMSE by the true value, expresses the error as the "ratio of divergence" from the true value.

$$RRMSE(\hat{\theta}) = \frac{RMSE(\hat{\theta})}{\theta}$$

Therefore, this method was dropped from the list of possible alternatives[21]. As the levels of RRMSEs for "industry-size" and the "10 closest enterprises" did not differ greatly, they were compared with the current method.

Then, the degrees of accuracy of the previous value imputation and the growth rate imputation, for both the "industry-size" and the "10 closest enterprises," were compared (Exhibit 3-2) [22] . With respect to the manufacturing industries, in fiscal 2009, when the rate of change from the previous year was larger than in other fiscal years because of the Lehman shock (Chart 6-1), the accuracy of the growth rate imputation was higher than that of the previous value imputation across the enterprise sizes, the greatest difference of RRMSEs being 9.5 percentage points for large enterprises. For the period other than fiscal 2009, the accuracy of the growth rate imputation was found to be higher than that of the previous value imputation for large and medium-sized enterprises. For small enterprises, the growth rate imputation brought about less accurate results. The average difference between the two methods was, however, only 0.2 percentage points. With respect to the non-manufacturing industries, for fiscal 2009, for medium-sized enterprises, and small enterprises when the "10 closest enterprises" was used, the growth rate imputation brought about more accurate results compared with the previous value imputation. For large enterprises, and small enterprises when the "industry-size" was used, the accuracy of the growth rate imputation was lower than from that of the previous value imputation. The difference between the two methods was,

---

[21] For "division by the number of employees" and the "10 closest enterprises," there were cases in which the value of the sum of the previous fiscal year's fixed investment of the respondent enterprises in the group was zero, making it impossible to obtain the growth rate. In such cases, the previous value imputation was applied.

[22] In this study, the period from fiscal 2004 to fiscal 2010 was divided into two periods based on the degree of change in the economic conditions. They are the "period 1" in which the rate of change from the previous year was larger than in other fiscal years and the rest as the "period 2." Then, the authors tried to verify that the accuracy was improved by the application of the alternative methods in the "period 1," while the accuracy of the alternative methods was higher than or more or less the same as that of the current method in the "period 2."

If the extent of improvement in accuracy is evaluated without dividing the period and on the basis of the average for fiscal 2004 through fiscal 2010, this would imply that there is a large economic change (of the size of the Lehman shock) as often as every seven years, which would result in the overestimation of the accuracy improvement effect of the alternative methods.

however, as small as 0.4 percentage points at the maximum. For the years other than fiscal 2009, the growth rate imputation led to less accurate results than the previous value imputation, though the difference between the two was, at most, 0.3 percentage points.

In summary, the alternative methods enabled more accurate imputation than the current method especially when applied to the manufacturing industries, for large ones among others, during the period of sharp changes in the economic conditions. Even when the alternative methods showed a poorer performance, the difference against the current method was small. Therefore, it is concluded that for fixed investment, the growth rate imputation including "industry-size" and the "10 closest enterprises" result in higher than or almost the same degree of accuracy as the current imputation method.

(Sales)

With respect to sales, a comparison of the accuracy of the three growth rate imputation methods (Exhibit 2-1) revealed that the RRMSEs tended to be larger for the "division by the number of employees" than for the other two methods (Exhibit 4-1). Thus, it was excluded from the list of the possible alternatives. A comparison of the other two growth rate imputation methods, namely "industry-size" and the "10 closest enterprises," showed that there was no significant difference in their RRMSEs. Thus, they were compared with the current imputation method.

The previous value imputation method was then compared with the growth rate imputation methods including "industry-size" and the "10 closest enterprises" in terms of accuracy of imputation (Exhibit 4-2). For both manufacturing and non-manufacturing industries, the growth rate imputation, on average, brought about more accurate results compared with the previous value imputation for all sizes of enterprises and for all years including 2009, when the rate of year-on-year change was larger than in other years on account of the Lehman shock (Exhibit 6-2).

In summary, the alternative methods made possible more accurate

imputation than the current method regardless of industry, period and the size of enterprises. Therefore, it is concluded that for sales, the growth rate imputation including "industry-size" and the "10 closest enterprises" result in a higher accuracy than the current method.

(Current profits)

With respect to current profits, the accuracy of the three ratio imputation methods (Exhibit 2-2) were compared (Exhibit 5-1). Because the "year before the missing value" showed cases in which the RRMSEs were clearly large (e.g., large enterprises in manufacturing industries in fiscal 2008), it was excluded from the possible alternatives. A comparison of the RRMSEs obtained from the "average response" and the "margin of increase" gave no clear indication as to which method would render better results than the other. Therefore, both of these methods were compared with the current method.

Then, the accuracy of the previous value imputation was compared with those of the ratio imputation methods (i.e., "average response" and the "margin of increase") (Exhibit 5-2). For manufacturing industries, both in fiscal 2008 and fiscal 2010 when the rate of year-on-year change was larger than in other years (Exhibit 6-3), the accuracy of the "average response" and the "margin of increase" was higher than that of the previous value imputation in all enterprise sizes. This tendency was clearer for large enterprises, for which both methods reduced RRMSEs by approximately 20 percentage points on average. For the years other than fiscal 2008 and fiscal 2010, the "margin of increase" brought about more accurate imputation for medium-sized and small enterprises, while for large enterprises the accuracy of those methods was slightly lower—by 0.1 percentage points on average—than the current method. The "average response" led to higher RRMSEs than the current method by 1.4 percentage points on average for large enterprises and by 0.1 percentage points on average for small enterprises. For non-manufacturing industries, the RRMSEs obtained from the ratio imputation methods were lower than the previous value imputation for all enterprise sizes when taking an average of fiscal 2008 and fiscal 2010. Between the two ratio imputation methods, the "margin of increase" brought

about better results. For the period other than fiscal 2008 and fiscal 2010, the "margin of increase" led to more accurate imputation for large and medium-sized enterprises, and to slightly less accurate results for small enterprises, an average difference of RRMSEs being 0.1 percentage points. For the same period, the "average response," brought about higher RRMSEs than the previous value imputation for all enterprise sizes, the average differences for each enterprise size being 0.8 percentage points at the maximum.

In summary, the alternative methods enabled more accurate imputation compared with the current method during the period of sharp changes in economic conditions. For other periods, although their results are worse than that of the current method in some cases, the difference between the "margin of increase" and the current method is small. Therefore, it is concluded that for current profits, the ratio imputation method—the "margin of increase" in specific—leads to a higher or about the same degree of accuracy as the current method.

(3) A summary of the appraisal of simulation results

In sum, the simulations suggest the following two points:

[i] For the period in which the rates of year-on-year change of the survey items are high, the alternative methods enable more accurate imputation than the current method by a large margin. Though in some cases the alternative methods lead to higher RRMSEs than the current method, the differences are small enough to allow us to judge that the degree of accuracy is about equal to that of the current method.

[ii] For periods other than those in which the rates of year-on-year change of the survey items are significant, the alternative methods can still be deemed to demonstrate a higher or about the same degree of accuracy as the current method.

Hence, it is concluded that for the three major survey items, there are alternative methods which produce a higher or about the same degree of accuracy as the imputation method now in use (Table below).

[Comparison of RRMSEs between the Simulations under the Current and Alternative Methods]

◎ ... Smaller RRMSEs for alternative methods for all enterprise sizes
○ ... About the same, even though there are cases in which alternative methods lead to higher RRMSEs

		Period in which the rate of year-on-year percent change rose due to Lehman shock Fixed investment and sales...FY09 Current profits...FY08 and FY10		Other periods	
Fixed investment Growth rate imputation [Industry-size] [10 closest enterprises]	Manufacturing	Smaller RRMSEs (in the case of alternative methods than in the case of the current method, the same shall apply hereinafter) for all sizes (−1.0 — −9.5%P).	◎	Slightly smaller RRMSEs for large and medium-sized (−0.0 — −0.3%P). Slightly larger RRMSEs for small enterprises (0.2%P).	○
	Non-manufacturing	・Smaller RRMSEs for medium-sized enterprises(−1.2%P). Slightly larger RRMSEs for large enterprises (0.3 — 0.4%P). Slightly larger RRMSEs for small enterprises (0.2%P [industry-size]) or smaller RRMSEs(−0.3%P [10 closest enterprises]).	○	Slightly larger RRMSEs for all sizes (0.1 — 0.3%P).	○
Sales Growth rate imputation [Industry-size] [10 closest enterprises.]	Manufacturing	Smaller RRMSEs for all sizes (−0.6 — −3.0%P).	◎	Smaller RRMSEs for all sizes (−0.1 — −1.1%P).	◎
	Non-manufacturing	Smaller RRMSEs for all sizes (−0.1 — −4.5%P).	◎	Smaller RRMSEs for all sizes (−0.0 — −0.4%P).	◎
Current profits Ratio imputation [Margin of increase]	Manufacturing	Smaller RRMSEs for all sizes (−0.9 — −20.3%P).	◎	Smaller RRMSEs for medium and small enterprises (−0.2 — −0.8%P). Slightly larger RRMSEs for large enterprises (0.1%P).	○
	Non-manufacturing	Smaller RRMSEs for all sizes (−0.1 — −1.5%P).	◎	Smaller RRMSEs for large and medium (−0.1 — −0.4%P). Slightly larger RRMSE for small enterprises (0.1%P).	○

There are alternatives whose “accuracy is higher or about the same as the method now in use.”

These conclusions are in line with what Utsunomiya and Sonoda (2001) suggested: the difference in the accuracy of the growth rate imputation and the previous value imputation is significant during a period of sharp changes in economic conditions, while it is small in a normal period.

(4) Impact of the difference in imputation methods on the aggregates

Based on these results, the impact of applying the alternative imputation methods on the aggregate figures will now be examined. Specifically, for large enterprises of both manufacturing and non-manufacturing industries, the figures for the projection for the new fiscal year were computed using the alternative imputation method on the three major survey items[23], and were compared with the original figures based on the current previous value imputation method (Exhibit 7). It reveals that in the survey taken when the rate of change from the previous year was high—for example, the March 2009 survey immediately after the Lehman shock—the aggregate figures based on the previous value imputation now in use and those based on the alternative methods differ by a large margin, especially for manufacturing industries. It is also observed that the aggregate figure based on the alternative methods showed a larger rate of change from the previous year.

A closer look at the data for fixed investment by large manufacturing enterprises in March 2009 survey indicates that both the absolute value of projection for fiscal 2009 and the missing value rates for the survey stood at approximately 15 percent, which were higher than the average of less than 10 percent and the average of about 10 percent, respectively for fiscal 2004 through fiscal 2008 (Exhibit 8-[ii]). This suggest that the large difference in the aggregate figures can be attributed to the large absolute value of projection itself combined with the high missing value rates (Exhibit 8-[iii]). The differences in aggregate figures grew smaller in June and later surveys as the missing value rates declined.

On the other hand, Exhibit 8-[iii] shows that there was no significant difference between the growth rate imputation and the previous value imputation for the years other than fiscal 2009. Thus, it is concluded that when the absolute value of projection and the missing value rates are at their normal levels, the growth rate imputation brings about little difference from the previous value imputation.

---

[23] The projection for the new fiscal year is obtained from March survey, when the survey for the data is conducted for the first time. The figures are expressed as year-on-year percent change.

5. Conclusions

Recognizing the challenges of the missing value imputation method now in use for the *Tankan*, this study has examined a number of alternative methods deemed appropriate for the major quantitative survey items such as fixed investment, sales and current profits.

The accuracy of these alternative methods was examined in comparison with the previous value imputation now in use based on a number of simulations. It has found that for the three major survey items of fixed investment, sales and current profits, there are alternative methods which produce a higher or about the same degree of accuracy as the current imputation method.

Two things should be noted for the conclusions of the study: this study has chosen the alternative methods which are feasible under the current survey practices of the *Tankan*; and the simulation this study has conducted can lead to different conclusions depending on the future economic conditions as well as the profile of missing values.

This study has concluded that there are alternative imputation methods which can materialize a higher or about the same degree of accuracy as that of the current method. However, this does not necessarily guarantee that the alternative methods are more appropriate for all survey items, periods, industries and the sizes of enterprises. Hence, in discussing the application of the alternative methods to the *Tankan*, the scope of application needs to be considered carefully.

In a phase of drastic changes of economic conditions, the choice of a missing value imputation method could greatly affect the estimated values. Therefore, improving missing value imputation methods in business surveys is a critical theme in compiling statistics. For the *Tankan*, it is important to continue studying more appropriate missing value imputation methods taking account of the results of this study.

References

Brown, I. and Terri, L.C. (2009), "Evaluation of Alternative Imputation Methods for the Public Libraries Survey (PLS)". Governments Division Report Series, Research Report #2009-4.

Eric, S.N. (1998), "Imputation: Methods, Simulation Experiments and Practical Examples". International Statistical Review, 66, 2, 157-180.

Ishikawa, A.,; Endo, S. and Shiratori, T. (2010), "Treatment of Outliers in Business Surveys: The Case of Short-Term Economic Survey of Enterprises in Japan (Tankan)", The Bank of Japan Working Paper.

Iwasaki, M. (2002), *Fukanzen Deta no Tokei Kaiseki* (*Statistical Analysis of Incomplete Data*), EconomistSha, Inc. (in Japanese)

Kataoka, M. (2010), "A Primer on the Tankan —Characteristics and Peculiarities of Major Items—", The Bank of Japan Review.

Lenka, M. (1995), "Imputation VS Reweighting for Total Nonresponse in a Business Survey". Proceedings of the Survey Methods Section, 69-74.

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

Mitsubishi Research Institute, Inc. (2010), "Tokutei Sabisu Sangyo Jittai Chosa to ni Okeru Suikei Shuho no Kakuritsu ni Kansuru Chosa Kenkyu" Chosa Hokokusho (Report on "A Study for Estimation Methods for Survey of Selected Service Industries, etc.") (in Japanese)

Rebecca, R.A. and Roderick, J.L. (2009). "The Use of Sample Weights in Hot Deck Imputation". Journal of Official Statistics, 25,1, 21-36.

Takezawa, K.. (2007), *Minna no Tame no Non Parametorikku Kaiki (Jo)*, (*Non-Parametric Regression for Everyone, Vol. 1*), Yoshioka-Shoten. (in Japanese)

Ton, W.; Jeroen, P. and Sander, S. (2010), Handbook of Statistical Data Editing and Imputation, Wiley.

Tsuchiya, T (2009), *Gaisetsu: Hyohon Chosa Ho* (*Sampling Survey Methods*), Asakura Shoten. (in Japanese)

Utsunomiya, K. and Sonoda, K. (2002), "Methodology for Handling Missing Values In TANKAN", The Bank of Japan Working Paper.

(Exhibit 1 - 1)

Survey Items of the *Tankan*

Judgement items		(1) Business Conditions (2) Domestic Supply & Demand Conditions for Products and Services (3) Overseas Supply & Demand Conditions for Products (4) Inventory Level of Finished Goods & Merchandise (5) Wholesalers' Inventory Level (6) Production Capacity (7) Employment Conditions (8) Financial Position (9) Lending Attitude of Financial institutions (10) Change in Interest Rate on Loans (11) Conditions for CP Issuance (12) Change in Output Prices (13) Change in Input Prices
	"Forecast Judgement" on quarterly data	(1) Interest-bearing Debt Outstanding (2) Level of Liquidity (3) Number of Employees
Quantitative items	Annual Projections	(1) Sales (2) Exports (3) Exchange Rates for Exports (4) Material costs (5) Personnel Expenses (6) Depreciation Expenses (7) Operating Profits (8) Financial Income (9) Financial Expenses (10) Current Profits (11) Net Income (12) Fixed Investment (13) Land Purchasing Expenses (14) Software Investment 3 Major Items (Examined in this study)
	Quarterly Data	(1) Total Liabilities (2) Loans from Financial institutions (3) Commercial Paper (4) Corporate Bonds (5) Total Assets (6) Cash & Deposits (7) Securities Listed as Liquid Assets (8) Securities Listed as Fixed Assets (9) Number of Employees (10) Part-time Workers
	Number of New Graduates Hired	Number of New Graduates Hired

(Exhibit 1 - 2)

Frequently Used Economic Statistics (Keidanren survey)

	Name of Statistics	Index	
1	The *Tankan* (Business Conditions)	78.2	← Judgement survey
2	Quarterly Estimates of GDP	75.8	
3	Consumer Price Index	70.7	
4	Labour Force Survey	66.9	
5	Indexes of Business Conditions	65.3	
6	The *Tankan* (Fixed Investment Projections, etc.)	65.1	← Quantitative item
7	Corporate Goods Price Index	61.5	
8	Indices of Industrial Production, Shipment and Inventory	61.0	
9	Trade Statistics	59.6	
10	The *Tankan* (Sales and Profit Projections)	59.1	← Quantitative item

Note: The survey was taken in June-July 2004, Covering 234 enterprises, including the members of Keidanren which were also its permanent directors, and other major member enterprises, including think tanks. The survey asked the members to evaluate 72 statistics on business conditions as "Always use it when released (3 points)," "Sometimes use it (2 points)," "Have used it (1 point)," "Never used it (0 point)." The points were aggregated for each statistics and the average value as a percentage of the maximum possible point (3) is calculated as an index.

Source: Nippon Keidanren (Japan Federation of Economic Organizations), "*Tokei no Riyo Kakudai ni Mukete*" (For Increased Use of Statistics), (2004)

(Exhibit 2 - 1)

Formulas for Growth Rate Imputation

(Applied to fixed investment and sales)

Name	Imputed value
[i] Growth rate imputation [Industry-size]	(Previous term's value of the non-respondent) × (Average growth rate of the respondents in the [Industry-size] group to which the non-respondent belongs) $\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{y_i^{(t-1)}}_{\text{Previous term's value of the non-respondent}} \times \underbrace{\frac{\sum y_k^{(t)}}{\sum y_k^{(t-1)}}}_{\text{Average growth rate between the previous term and this term of the respondents in the (industry-size) group to which the non-respondent belongs}}$
[ii] Growth rate imputation [Division by employees]	(Previous term's value of the non-respondent)× (Average growth rate of the respondents in the [Division by employees] group to which the non-respondent belongs) $\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{y_i^{(t-1)}}_{\text{Previous term's value of the non-respondent}} \times \underbrace{\frac{\sum y_k^{(t)}}{\sum y_k^{(t-1)}}}_{\text{Average growth rate between the previous term and this term of the respondents in the (division by employees) group to which the non-respondent belongs}}$
[iii] Growth rate imputation [10 closest enterprises]	(Previous term's value of the non-respondent) × (Average growth rate of the respondents in the [10 closest enterprises] group in terms of fixed investment or sales in the [Industry-size] group to which the non-respondent belongs) $\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{y_i^{(t-1)}}_{\text{Previous term's value of the non-respondent}} \times \underbrace{\frac{\sum y_k^{(t)}}{\sum y_k^{(t-1)}}}_{\text{Average growth rate between the previous term and this term of the respondents in the (10 closest enterprises) group to which the non-respondent belongs}}$ * Enterprises are listed according to the size of fixed investment or sales in a descending order and the 10 enterprises (5 above and 5 below) closest to the non-respondent were chosen.

* For the [industry-size] group, the weight (= the number of enterprises in the population/the number of respondents) for each stratum of the population to which the respondent belongs was established to obtain the growth rate by multiplying the value reported by individual respondent by this figure.

(Exhibit 2 - 2)

Formulas for Ratio Imputation (Applied to current profits)

Name	Imputed value
[i] Ratio imputation [Previous year]	(This term's sales of the non-respondent) × ("Current profits/sales" ratio of the non-respondent in the previous term... a) $\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{x_i^{(t)}}_{\text{This term's value for another item reported by the non-respondent}} \times \underbrace{\frac{y_i^{(t-1)}}{x_i^{(t-1)}}}_{\text{Previous term's ratio between the items of the non-respondent}}$
[ii] Ratio imputation [Average response]	(This term's sales of the non-respondent) × (This term's average "current profits/sales" ratio of the respondents in the [industry-size] group to which the non-respondent belongs) $\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{x_i^{(t)}}_{\text{This term's value for another item reported by the non-respondent}} \times \underbrace{\frac{\sum y_k^{(t)}}{\sum x_k^{(t)}}}_{\text{This term's average ratio between the items of the (industry-size) group to which the non-respondent belongs*}}$ Approximate value of "this term's ratio between the items of the non-respondent"
[iii] Ratio imputation [Margin of increase]	(This fiscal year's sales of the non-respondent) × (a in [i] in this table + Average margin of increase in "current profits/sales" ratio of the [industry-size] group to which the non-respondent belongs) $\underbrace{\tilde{y}_i^{(t)}}_{\text{Imputed value}} = \underbrace{x_i^{(t)}}_{\text{This term's value for another item reported by the non-respondent}} \times \underbrace{\left\{\frac{y_i^{(t-1)}}{x_i^{(t-1)}}\right.}_{\text{Previous term's ratio between the items of the non-respondent}} + \underbrace{\left.\left(\frac{\sum y_k^{(t)}}{\sum x_k^{(t)}} - \frac{\sum y_k^{(t-1)}}{\sum x_k^{(t-1)}}\right)\right\}}_{\text{Average margin of change (margin of increase) in the ratio between the items between the previous term and this term of the (industry-size) group to which the non-respondent belongs}}$ Approximate value of "this term's ratio between the items of the non-respondent"

* For [average response] and [margin of increase], the weight (= the number of enterprises in the population / the number of respondents) for each stratum of the population to which the respondent belongs was established, to obtain the ratio between items by multiplying the value reported by individual company by this figure.

(Exhibit 3 - 1)

A Comparison of RRMSEs for Fixed Investment [i]

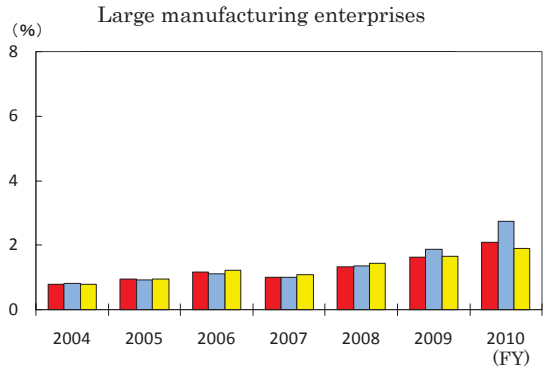
... Growth rate imputation [industry-size]

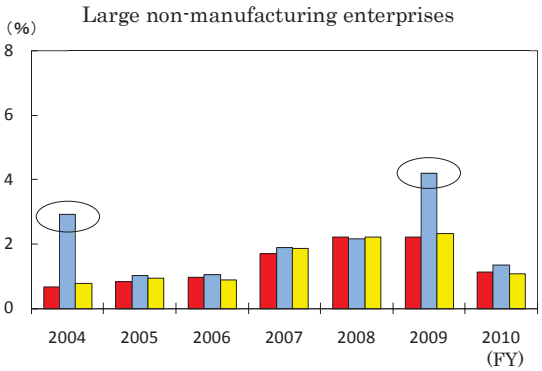
... Growth rate imputation [Division by employees]

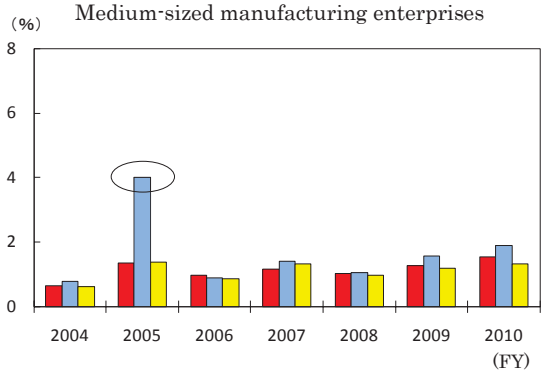
... Growth rate imputation [10 closest enterprises]

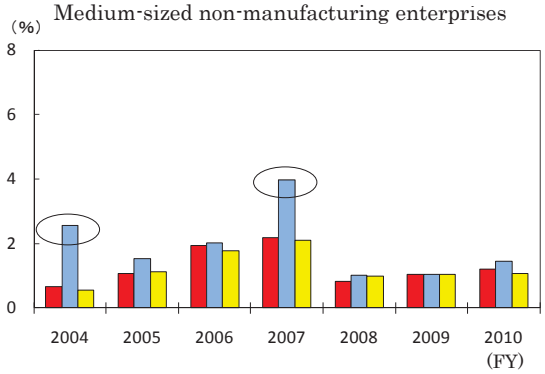
are excluded because RRMSEs are clearly large for some fiscal years.

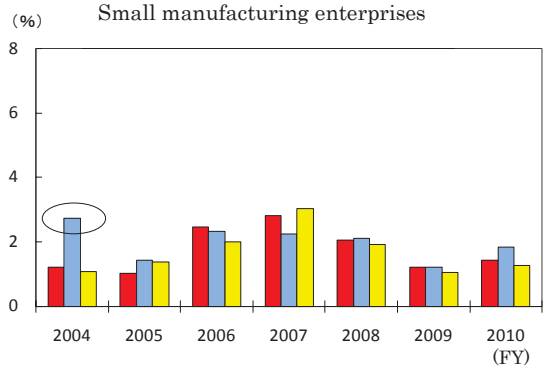
& show no significant difference in RRMSEs.

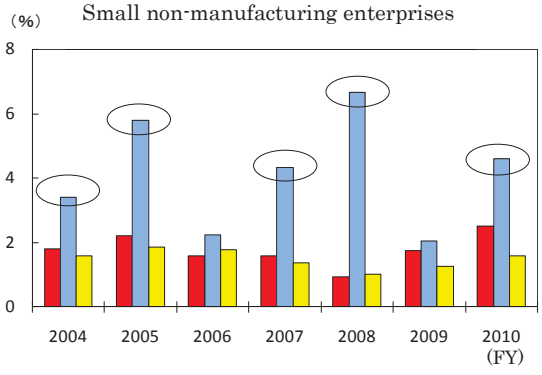












(Exhibit 3 - 2)

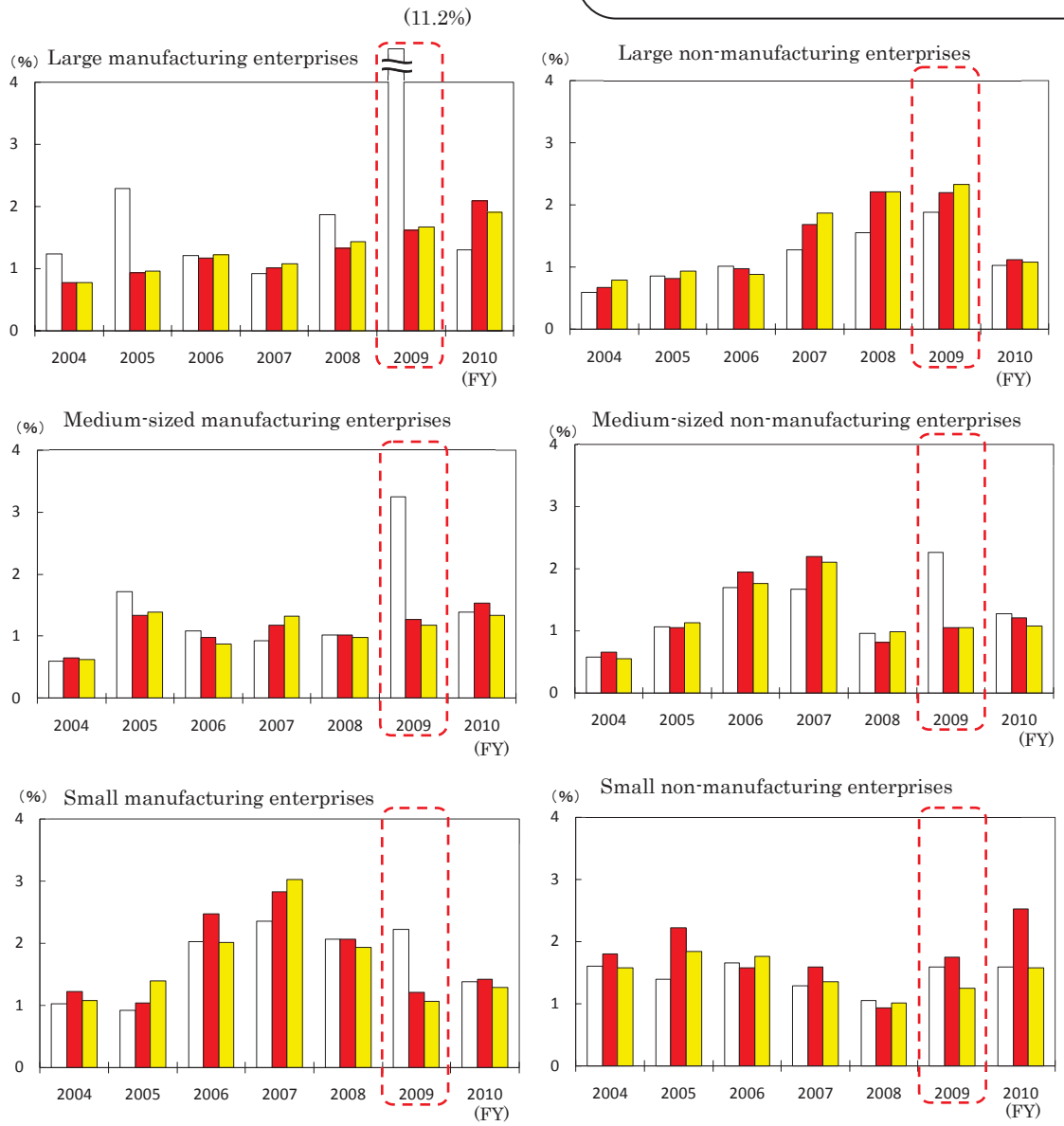
A Comparison of RRMSEs for Fixed Investment [ii]

... Previous value imputation ... Growth rate imputation [industry-size]

... Growth rate imputation [10 closest enterprises]

Manufacturing…For FY09, the RRMSEs of and are clearly small for enterprises in all sizes. For other fiscal years, the RRMSEs of and are small on average for large and medium-sized enterprises. The RRMSEs are worse for small enterprises, but the margin of deterioration is 0.2 percentage points.

Non-manufacturing: For FY09, the RRMSEs of and for medium-sized enterprises and of for small enterprises are small. The RRMSEs are worse for others, but the margin of deterioration is at most 0.4 percentage points. For other fiscal years, the RRMSEs of and are worse, but the margin of deterioration is at most 0.3 percentage points.



Degree of Improvement of RRMSEs by the Application of Alternative Methods*

Growth Rate Imputation (industry-size)

(% points)

		FY 09	Average for other FYs	Average for FY04–FY10
Manufacturing	Large	−3.0	−1.1	−1.4
	Medium	−0.9	−0.2	−0.3
	Small	−0.6	−0.1	−0.1
Non-manufacturing	Large	−4.4	−0.4	−0.9
	Medium	−0.3	−0.0	−0.1
	Small	−0.1	−0.0	−0.0

Growth Rate Imputation (10 closest enterprises)

(% points)

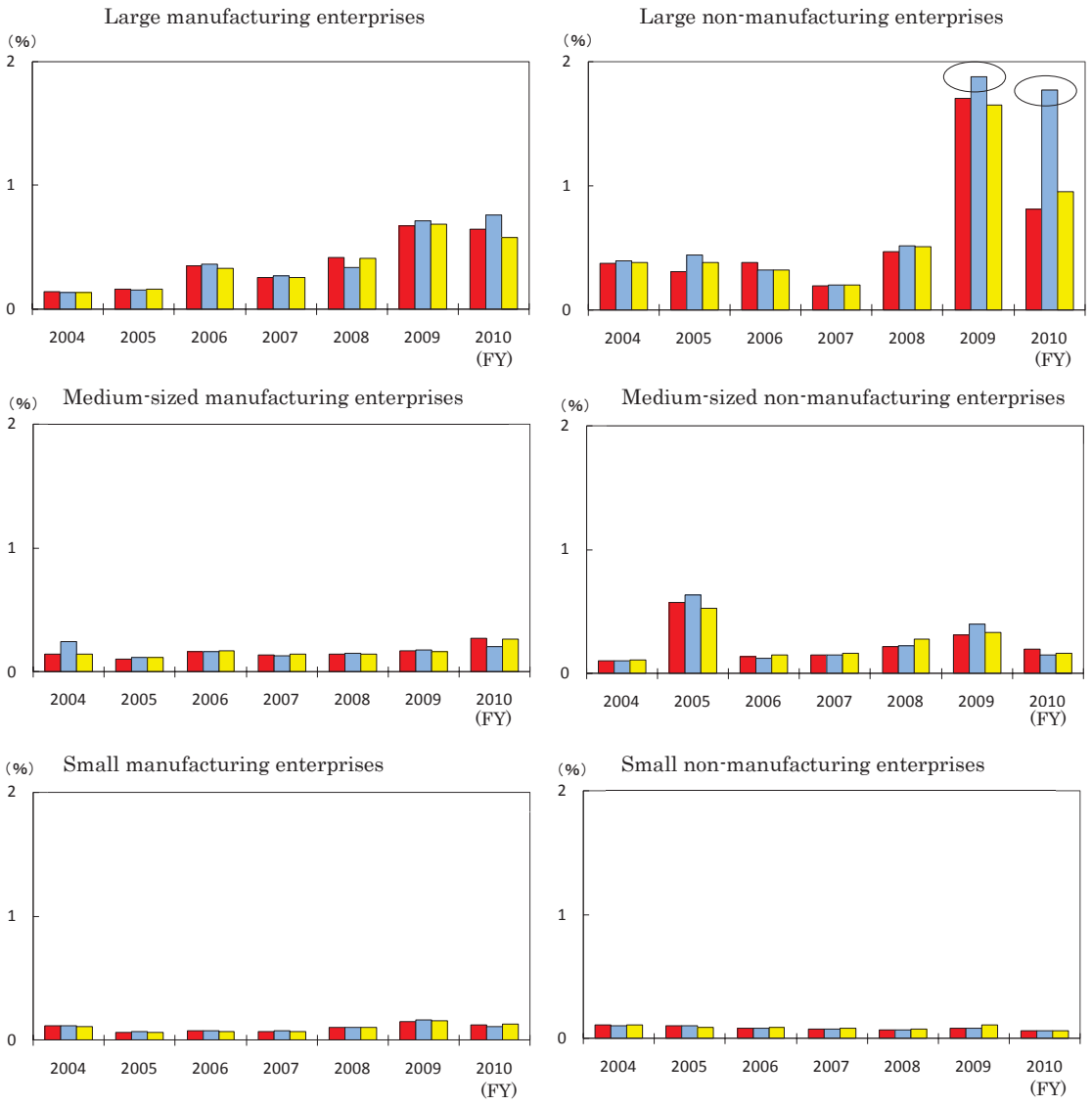
		FY 09	Average for other FYs	Average for FY04–FY10
Manufacturing	Large	−3.0	−1.1	−1.4
	Medium	−0.9	−0.2	−0.3
	Small	−0.6	−0.1	−0.1
Non-manufacturing	Large	−4.5	−0.3	−0.9
	Medium	−0.3	−0.0	−0.0
	Small	−0.1	−0.0	−0.0

(Exhibit 4 - 1)

A Comparison of RRMSEs for Sales [i]

... Growth rate imputation [industry-size]
... Growth rate imputation [division by employees]
... Growth rate imputation [10 closest enterprises]

are excluded because the RRMSEs tend to be large.
& show no significant difference in the RRMSEs.



(Exhibit 4 - 2)

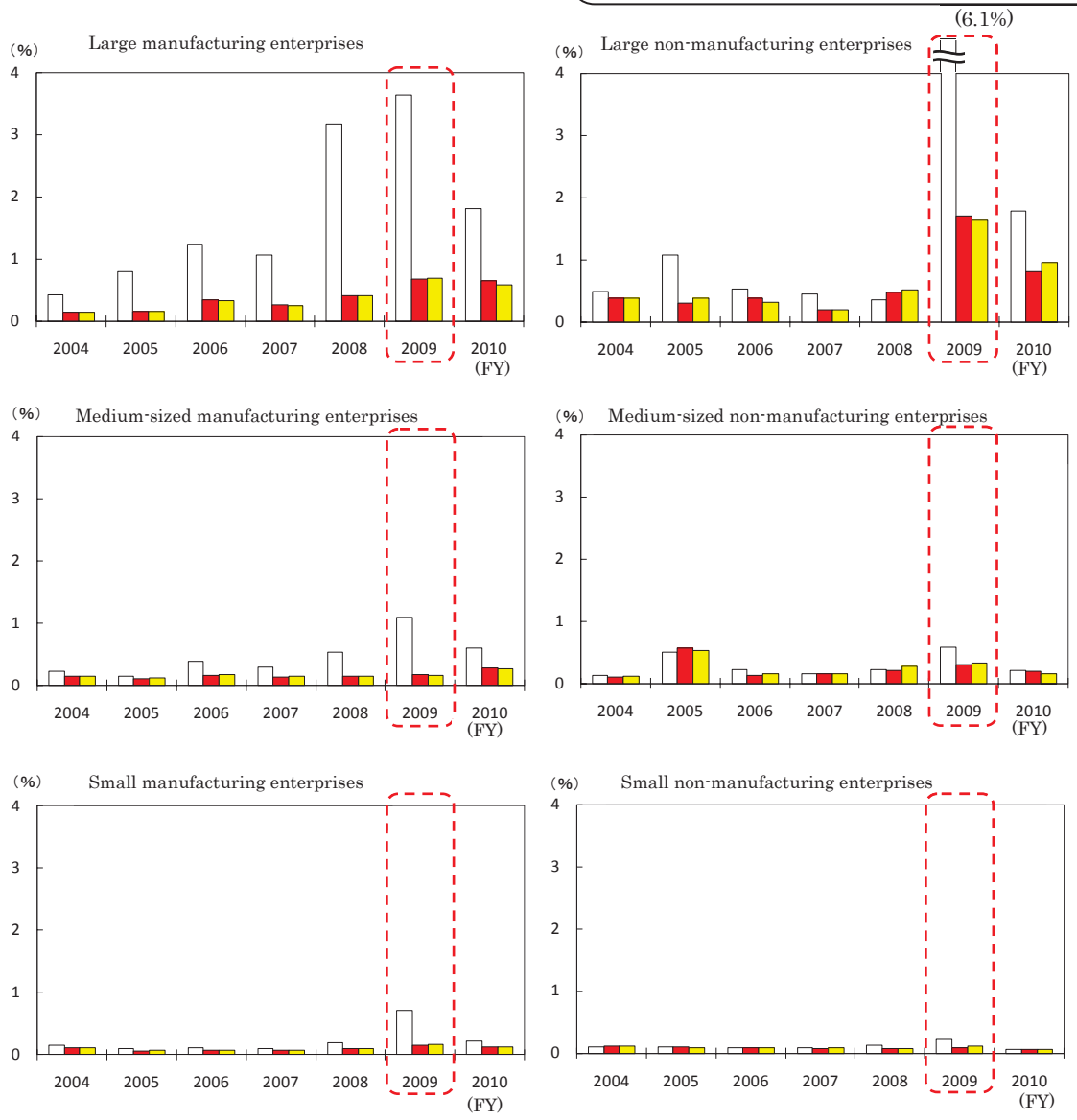
A Comparison of RRMSEs for Sales [ii]

... Previous value imputation ... Growth rate imputation [industry-size]

... Growth rate imputation [10 closest enterprises]

Manufacturing: In all fiscal years, the RRMSEs for and are small for all enterprise sizes.

Non-manufacturing: FY09, the RRMSEs for and are small for all enterprise sizes. For other fiscal years, the RRMSEs for and are small on average for all enterprise sizes.



Degree of Improvement of RRMSEs by the Application of Alternative Methods*

Growth Rate Imputation (industry-size) (% points)

		FY 09	Average for other FYs	Average for FY04–FY10
Manufacturing	Large	−3.0	−1.1	−1.4
	Medium	−0.9	−0.2	−0.3
	Small	−0.6	−0.1	−0.1
Non-manufacturing	Large	−4.4	−0.4	−0.9
	Medium	−0.3	−0.0	−0.1
	Small	−0.1	−0.0	−0.0

Growth Rate Imputation (10 closest enterprises) (% points)

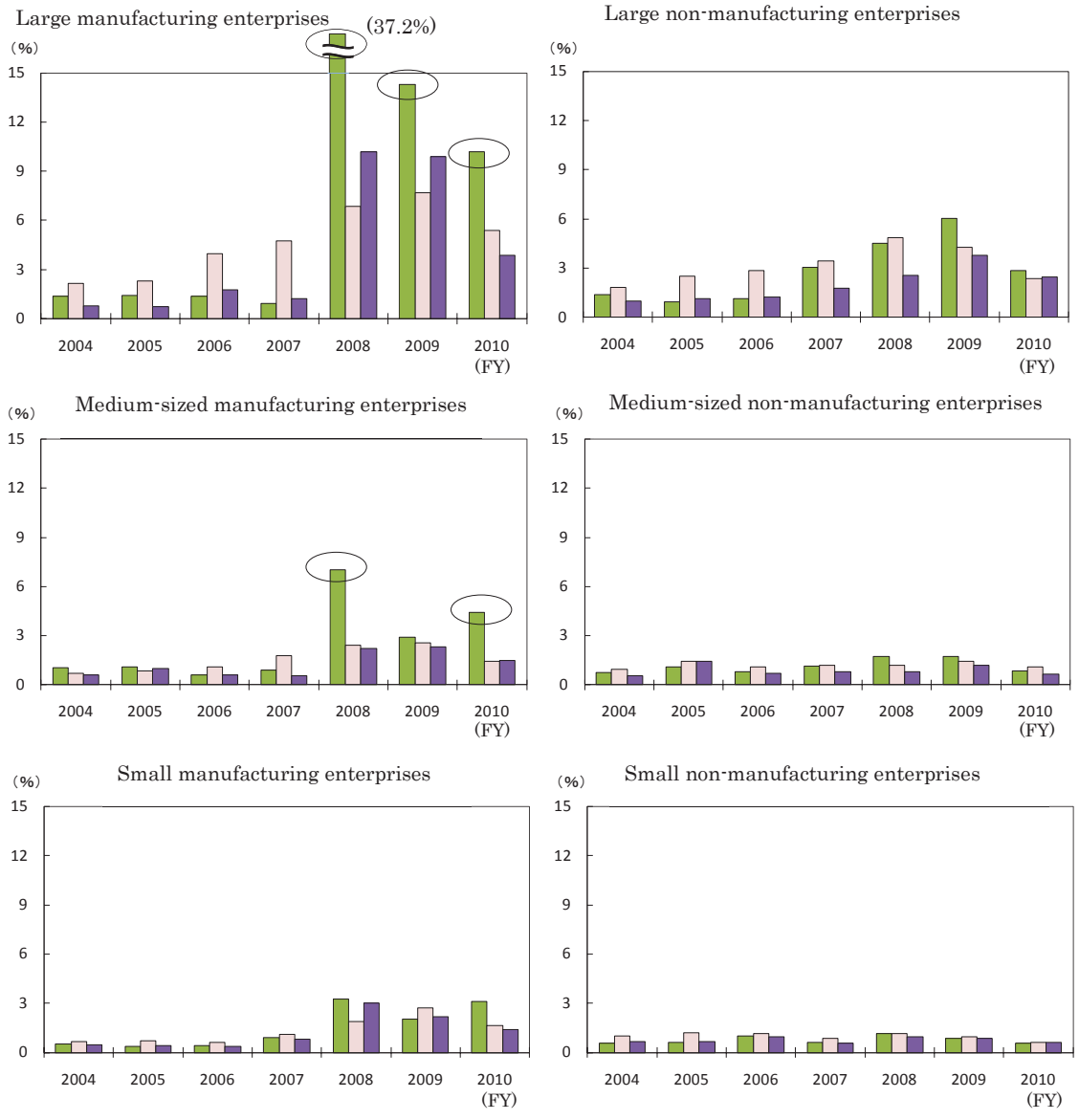
		FY 09	Average for other FYs	Average for FY04–FY10
Manufacturing	Large	−3.0	−1.1	−1.4
	Medium	−0.9	−0.2	−0.3
	Small	−0.6	−0.1	−0.1
Non-manufacturing	Large	−4.5	−0.3	−0.9
	Medium	−0.3	−0.0	−0.0
	Small	−0.1	−0.0	−0.0

(Exhibit 5 - 1)

A Comparison of RRMSEs for Current Profits [i]

... Ratio imputation [the year before missing value]
... Ratio imputation [average response]
... Ratio imputation [margin of increase]

are excluded because the RRMSEs are clearly large in some cases.
& are difficult to judge which is better.



(Exhibit 5 - 2)

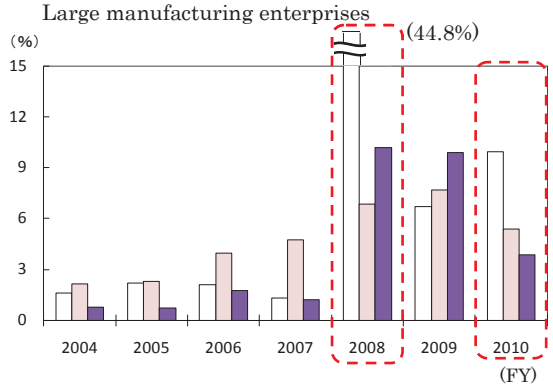
A Comparison of RRMSEs for Current Profits [ii]

... Previous value imputation ... Ratio imputation [average response]
... Ratio imputation [margin of increase]

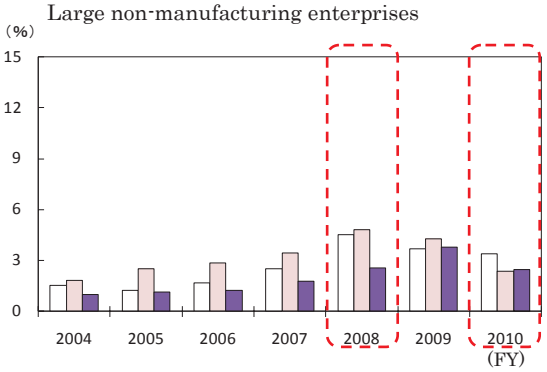
Manufacturing: The averages for FY08 and FY10 show that the RRMSEs for and are small for all enterprise sizes. For other years, the RRMSEs for are higher by at most 0.1 percentage points (1.4 percentage points for) than .

Non-manufacturing: The averages FY08 and FY 10 show that the RRMSEs for and are small for all enterprise sizes. For other years, the RRMSEs for are higher by at most 0.1 percentage points (0.8 percentage points for) than .

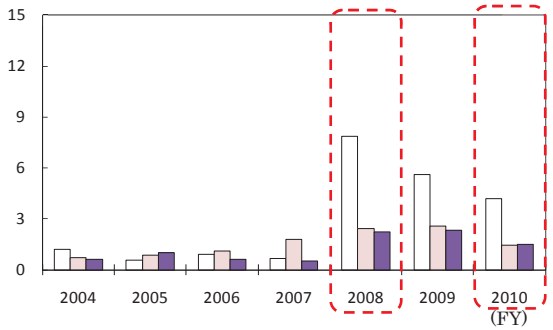
Large manufacturing enterprises



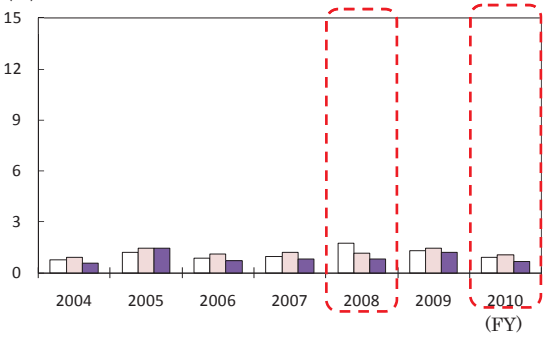
Large non-manufacturing enterprises



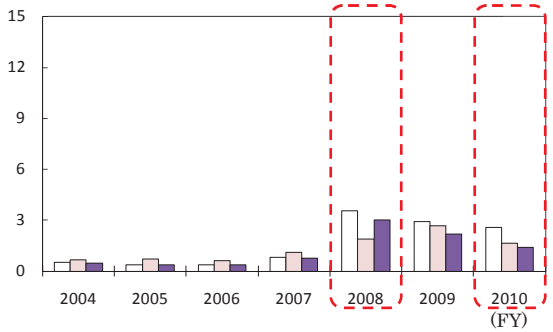
Medium-sized manufacturing enterprises



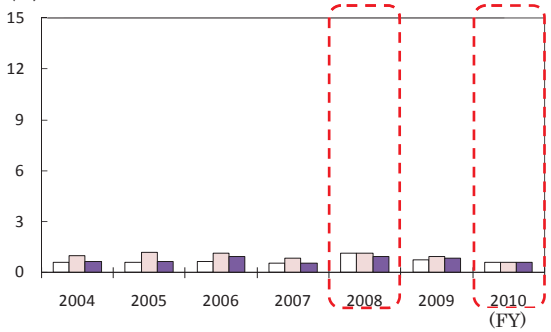
Medium-sized non-manufacturing enterprises



Small manufacturing enterprises



Small non- manufacturing enterprises



Degree of Improvement of RRMSEs by the Application of Alternative Methods*

Growth Rate Imputation (industry-size)

(% points)

		FY 09	Average for other FYs	Average for FY04–FY10
Manufacturing	Large	−21.2	1.4	−5.1
	Medium	−4.1	−0.4	−1.4
	Small	−1.3	0.1	−0.3
Non-manufacturing	Large	−0.4	0.8	0.5
	Medium	−0.2	0.2	0.1
	Small	−0.0	0.4	0.3

Growth Rate Imputation (10 closest enterprises)

(% points)

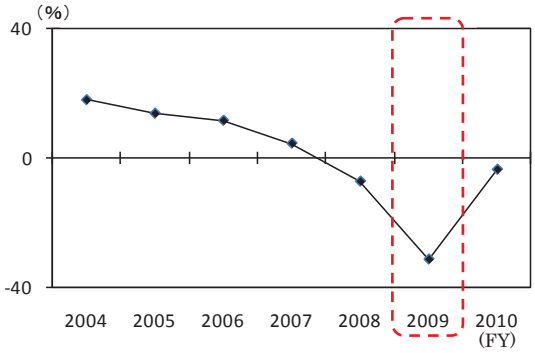
		FY 09	Average for other FYs	Average for FY04–FY10
Manufacturing	Large	−20.3	0.1	−5.7
	Medium	−4.2	−0.8	−1.7
	Small	−0.9	−0.2	−0.4
Non-manufacturing	Large	−1.5	−0.4	−0.7
	Medium	−0.6	−0.1	−0.2
	Small	−0.1	0.1	0.0

(Exhibit 6 - 1)

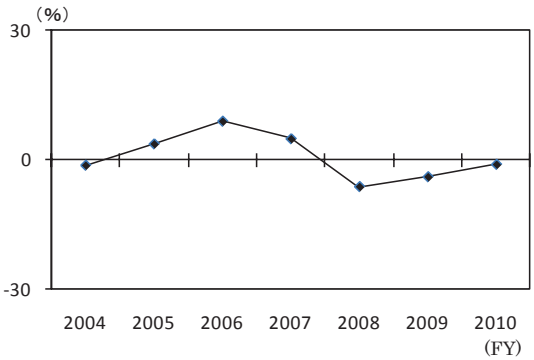
Rate of Year-to-year percent change in Fixed Investment

The rate of Year-to-year change in FY09 was higher than in other fiscal years in many enterprise sizes and industries.

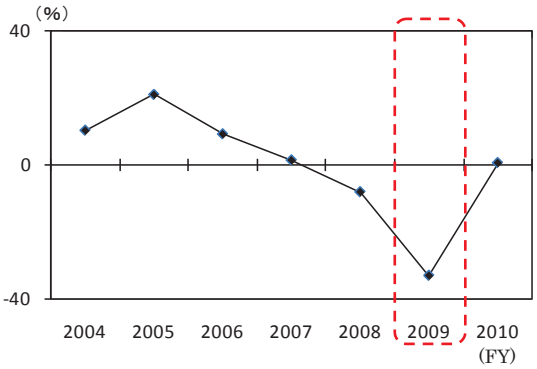
Large manufacturing enterprises



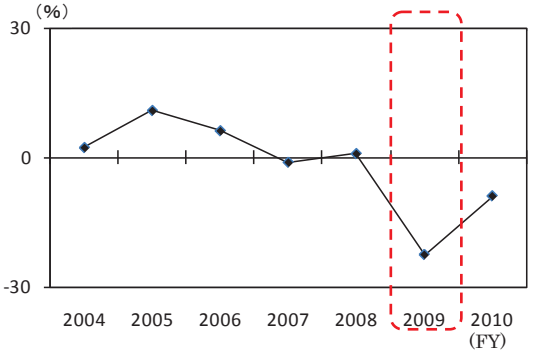
Large non-manufacturing enterprises



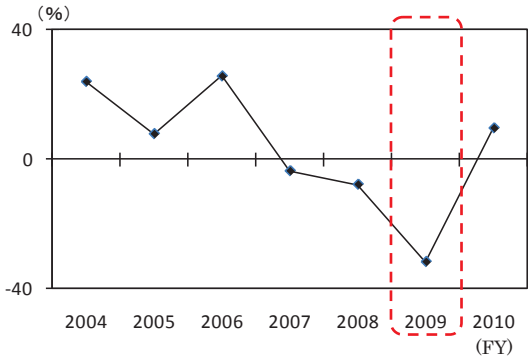
Medium-sized manufacturing enterprises



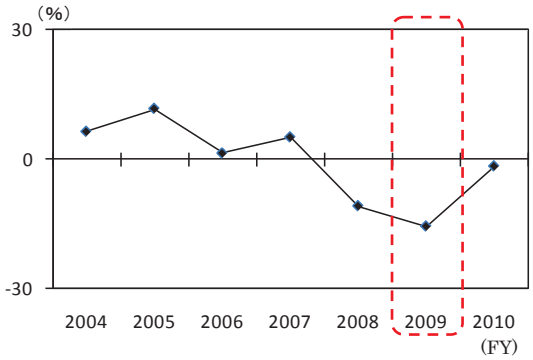
Medium-sized non-manufacturing enterprises



Small manufacturing enterprises



Small non- manufacturing enterprises

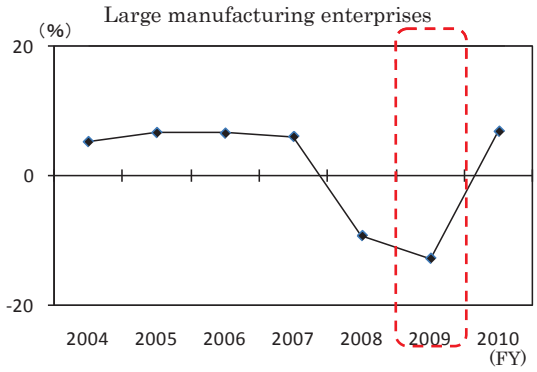


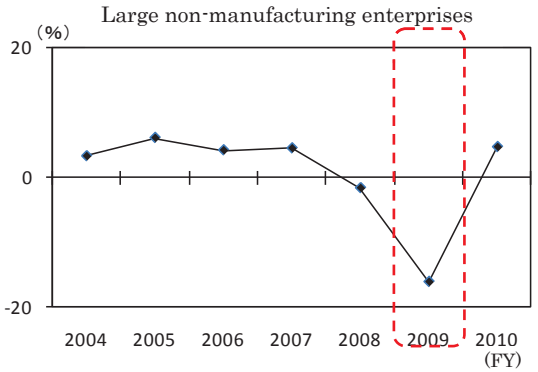
(Note)Actual data as compared with the previous year obtained in June surveys.

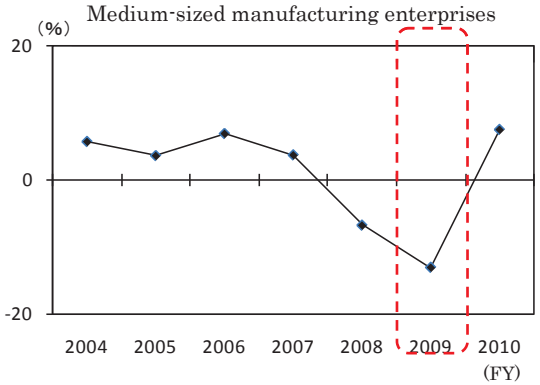
(Exhibit 6 - 2)

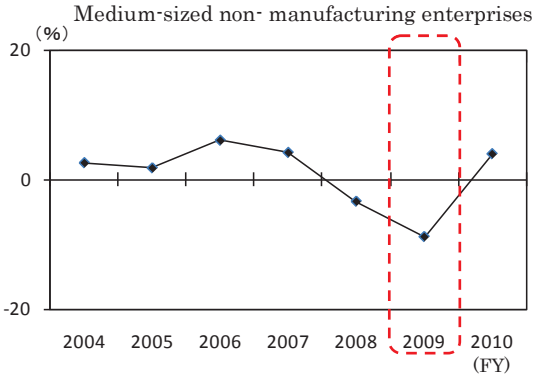
Rate of Year-to-year percent change in Sales

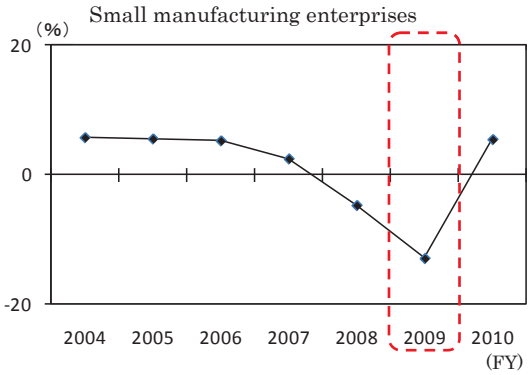
The rate of year-on-year change in FY09 is higher than in other fiscal years.

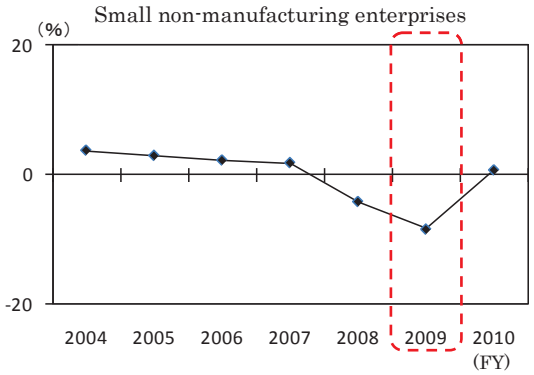












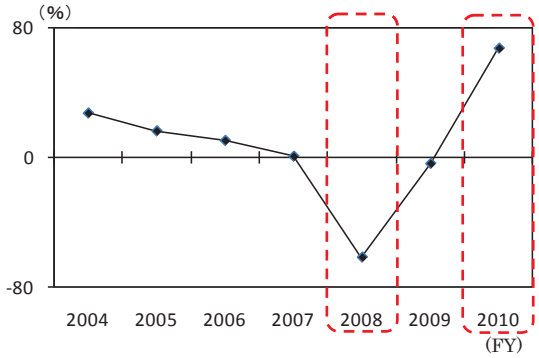
(Note)Actual data as compared with the previous year obtained in June surveys.

(Exhibit 6 - 3)

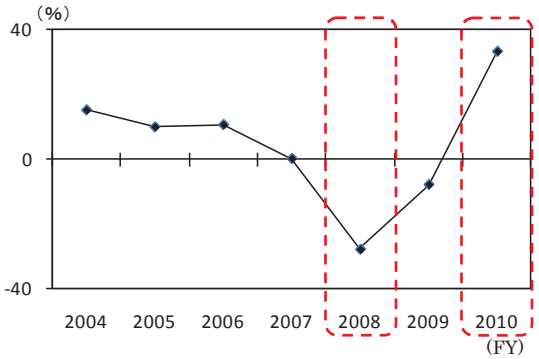
The Rate of Year-to-year percent change in Current Profits

The rates of year-on-year change in FY08 and FY10 are higher than in other fiscal years for many enterprise sizes and industries.

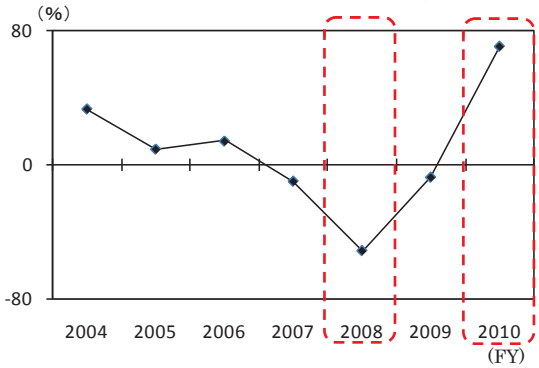
Large manufacturing enterprises



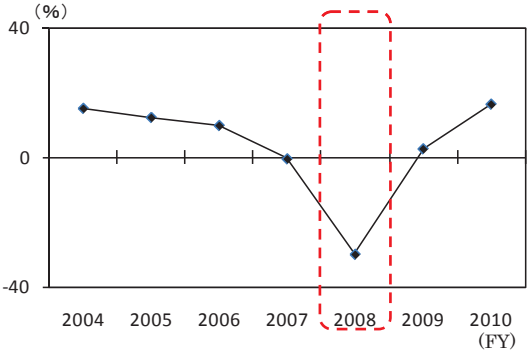
Large non-manufacturing enterprises



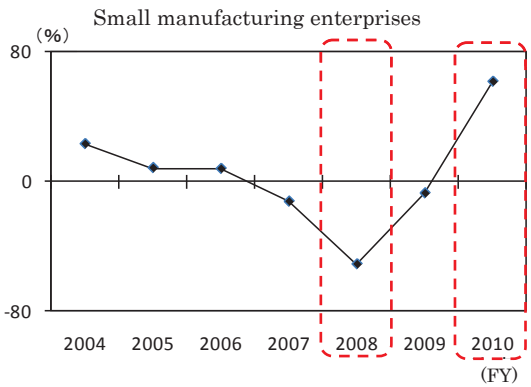
Medium-sized manufacturing enterprises



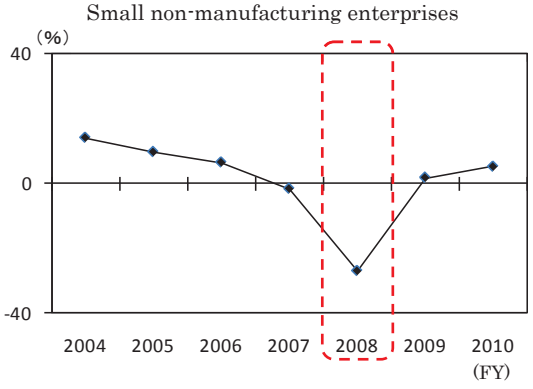
Medium- sized non-manufacturing enterprises



Small manufacturing enterprises



Small non-manufacturing enterprises



(Note) Actual data as compared with the previous year obtained in June surveys.

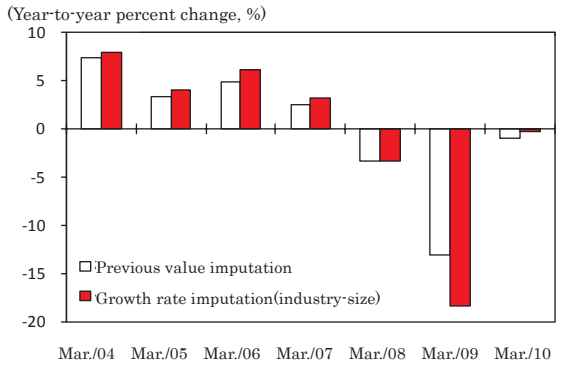
(Exhibit 7)

A Comparison of Current Method and Alternatives

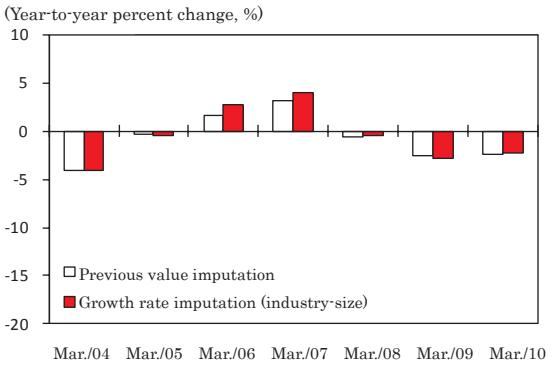
(Projection for New Fiscal Year obtained in March Surveys)

Fixed Investment

Large manufacturing enterprises

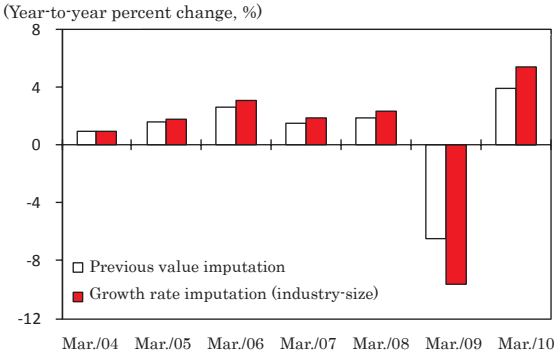


Large non-manufacturing enterprises

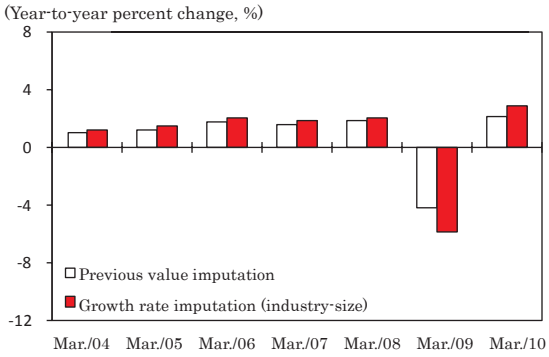


Sales

Large manufacturing enterprises

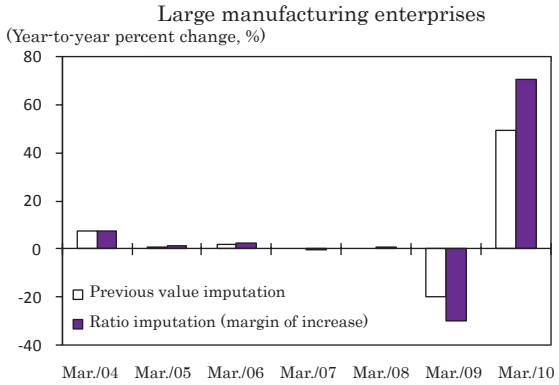


Large non-manufacturing enterprises

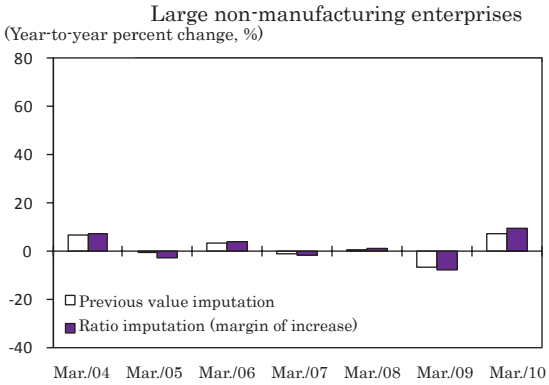


Current Profits

Large manufacturing enterprises



Large non-manufacturing enterprises

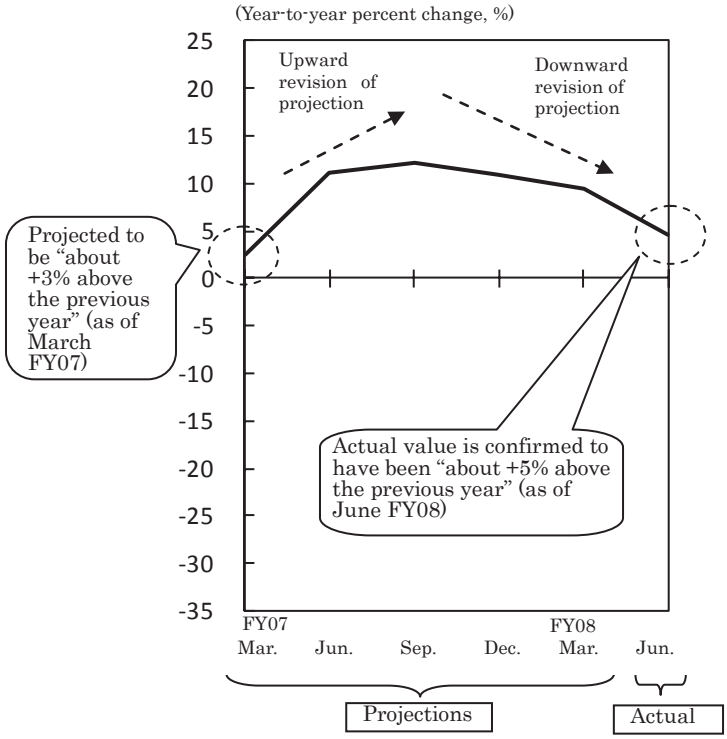


* The alternative method used is “growth rate imputation industry-size” for fixed investment and sales, and “ratio imputation margin of increase” for current profits.

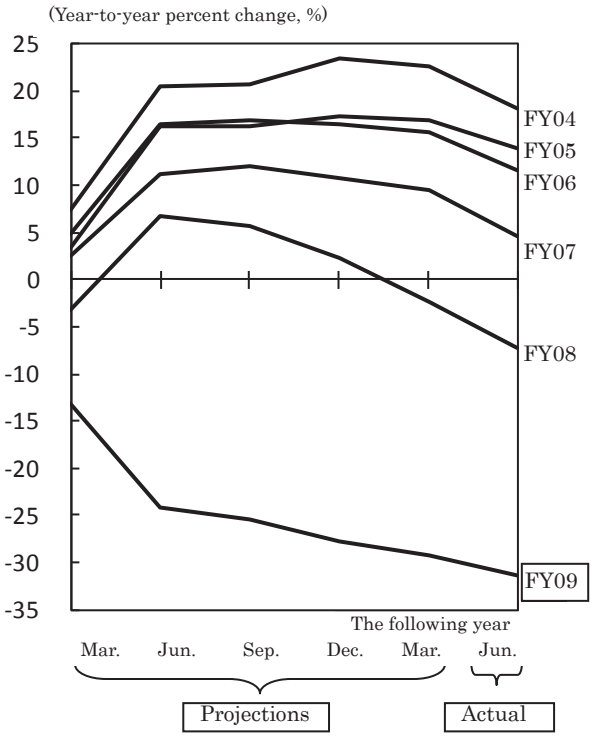
(Exhibit 8)

Pattern of Year-to-year Revisions of Fixed Investment (Large manufacturing)

[i] FY07 (Current method)



[ii] FY04-FY09 (Current method)



(Note) For example, for FY07, the "projection for FY07 as compared with the previous year" is compiled and released for the first time in March survey. Later, the "revised values of the projection" are released in surveys between June FY07 and March FY08, and the "actual value for FY07 as compared with the previous year" is released in June FY08, finalizing the value as compared with the previous year.

	Mar.	Jun.	Sep.	Dec.
Average missing value rate for FY04-FY08 (%)	10.2	2.7	2.0	1.7
Missing value rate for FY09 (%)	15.3	4.5	3.6	2.9

[iii] FY04-FY09 (Current and alternative methods)