(Appendix)

Treatment of Outliers

In the *Tankan* (Short-term Economic Survey of Enterprises in Japan), numerical items such as sales are estimated by "population estimates" -- a statistical method calculated by industry and size classifications. To enhance the precision of these statistical data, outliers from a population are treated in a specific rule in estimating population estimates. This rule is applied from the survey in December 2010. The rule is explained specifically in terms of "detection" and "treatment" of outliers as follows[1]:

1. Method of Outlier Detection

(1) y_{it}, the degree of influence of an enterprise on the rate of change from the previous year or the rate of revision from the previous survey of the six categories (Manufacturing・Nonmanufacturing sectors × Large・Medium・Small sized enterprises) is defined as follows:

$$y_{it} = \frac{\text{The margin of change in the weighted data}}{\left|\begin{array}{c}\text{The estimated value for the population on the six}- \\ \text{category base in the previous fiscal year (previous survey)}\end{array}\right|} \times 100$$

$$= \frac{w_i'(x_{it}-x_{it-1})}{|M_{jt-1}|} \times 100$$

x_{it} : The value reported by the i-th enterprise at time t

x_{it-1} : The value reported by the i-th enterprise at time t-1

w_i' : The weight of x_i after modification*

M_{jt-1} : The estimated value for the population in one of the six categories (j=1,....,6)

* To accurately measure "the margin of change in the weighted data," despite the change in weight between time t and time t-1, w_i' is modified to have the same number of sample enterprises between time t and t-1. (Therefore, if there is no response in either of the surveys (at two points in time), the missing value is replaced with the mean within each stratum to arrive at the same number of samples in both surveys.)

(2) z_{it}, the indicator of the degree of divergence of y_{it}, is defined as follows:

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

If the indicator of the degree of divergence, z_{it}, exceeds the standard (C=50), which is set in advance based on empirical analyses using past data, it is detected as an outlier.

d_j^{1} : 1 percentile[2] of y_{it} in the j-th category

d_j^{99} : 99 percentile of y_{it} in the j-th category

D_j : The unit of distance between d_j^{99} and d_j^{1} $(d_j^{99} - d_j^{1})$

---- Treatment of outliers is applied to the five items which are "sales," "current profits," "net income," "fixed investment," and "software investment." Additionally, "export" and "land investment" -- which are part of sales and fixed investment -- are accordingly treated as outliers only when sales and fixed investment are detected as outliers.

2. Method of Outlier Treatment

The detected outliers are going to be treated as missing values, since these outliers do not represent the population; it would not be appropriate to include such data. With the same rule as the current one, a value reported in the previous year or previous survey is substituted to a missing value $<x_{it} = x_{it-1}>$ to avoid influence of the outliers. Furthermore, once the data is regarded as an outlier, the treatment will be applied until the Bank of Japan reviews the sample enterprises for the *Tankan*.

---

[1] For details regarding outliers, please refer to "Treatment of Outliers in Business Surveys: The Case of Short-term Economic Survey of Enterprises in Japan (Tankan)" (July, 2010).

[2] The value which could be found at 1% from the smallest to the largest of observations.