Toward the Improvement of the Tankan Survey
— Imputing Missing Values —

2011.11.17
Takahiro Hirakawa
Economic Statistics Division
Research and Statistics Department
Bank of Japan
takahiro.hirakawa @ boj.or.jp

Notice:
Views expressed hereafter are those of the author and do not necessarily represent those of the Bank of Japan.
Today’s Topics

1. What is the Tankan survey?

2. Features of the Tankan survey

3. Improving imputation method
   — current method & problem
   — alternatives
   — comparing current method & alternatives by simulation

4. Conclusion
1. What is the Tankan survey?

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Providing an accurate picture of business conditions ⇒ Contributing to the appropriate implementation of monetary policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey Items</td>
<td></td>
</tr>
<tr>
<td>• Judgement survey (e.g., Business conditions index)</td>
<td></td>
</tr>
<tr>
<td>• Quantitative data (e.g., Sales, Fixed investment)</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
</tr>
<tr>
<td>• Sample enterprises: 10,910 (Sept. 2011)</td>
<td></td>
</tr>
<tr>
<td>• Population enterprises: 213,210</td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>Quarterly (March, June, September, December)</td>
</tr>
</tbody>
</table>
2. Features of the Tankan survey

① High response rate
- 98.8% (Sept. 2011)
  - keeping a high response rate (95.6%) even immediately after the Great East Japan Earthquake (March 2011)

② Long history
- The predecessor of the present Tankan started in 1957.
  - on the model of “Economic test” by the IFO Institute
- The present Tankan started in 1974.
  - registering the 150th survey in Sept. 2011

③ Statistical accuracy
- Regular revision of the sample enterprises
- Treatment of outliers
- Missing value imputation

④ Quick release
- Response collected in one month
- Released on the day after the end of the survey period

⑤ Various survey items
- Judgement survey (13 items)
- Quantitative data (25 items)

【Box 1】

【Box 2】

Virtuous Cycle

Understanding the significance of the Tankan survey

high credibility

high reputation & attention

Today’s main topic
### Judgment Survey (13 items)

- Business Conditions
- Domestic Supply and Demand Conditions for Products and Services
- Overseas Supply and Demand Conditions for Products
- Inventory Level of Finished Goods and Merchandise
- Wholesalers’ Inventory Level
- Production Capacity
- Employment Conditions
- Financial Position
- Lending Attitude of Financial institutions
- Change in Interest Rate on Loans
- Conditions for CP Insurance
- Change in Output Prices
- Change in Input Prices

### Quantitative Data (25 items)

- Aggregated into Diffusion Index
- Calculated from the amount of population estimates into year-to-year % changes and revision rates

### Annual Projections (14 items)

- Sales
- Exports
- Exchange rates for Exports
- Material Costs
- Personnel Expenses
- Depreciation Expenses
- Operating Profits
- Financial Income
- Financial expenses
- Current Profits
- Net Income
- Fixed Investment
- Land Purchasing Expenses
- Software Investment

### Quarterly Data (10 items)

- Total Liabilities
- Loans from Financial institutions
- Commercial Paper
- Corporate Bonds
- Total Assets
- Cash and Deposits
- Securities Listed as Liquid Assets
- Securities Listed as Fixed Assets
- Number of Employees
- Part-time workers
- Number of New Graduates Hired
<High reputation and attention to the Tankan survey>


<table>
<thead>
<tr>
<th>Name of statistics</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Tankan (Business Conditions)</td>
<td>78.2</td>
</tr>
<tr>
<td><strong>2</strong> Quarterly Estimates of GDP</td>
<td>75.8</td>
</tr>
<tr>
<td><strong>3</strong> Consumer Price Index</td>
<td>70.7</td>
</tr>
<tr>
<td><strong>4</strong> Labour Force Survey</td>
<td>66.9</td>
</tr>
<tr>
<td><strong>5</strong> Indexes of Business Conditions</td>
<td>65.3</td>
</tr>
<tr>
<td><strong>6</strong> Tankan (Fixed Investment Projections, etc.)</td>
<td>65.1</td>
</tr>
<tr>
<td><strong>7</strong> Corporate Goods Price Index</td>
<td>61.5</td>
</tr>
<tr>
<td><strong>8</strong> Indices of Industrial Production, Shipment and Inventory</td>
<td>61.0</td>
</tr>
<tr>
<td><strong>9</strong> Trade Statistics</td>
<td>59.6</td>
</tr>
<tr>
<td><strong>10</strong> Tankan (Sales and Current Profit Projections)</td>
<td>59.1</td>
</tr>
</tbody>
</table>

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 Muketē" (For Increased Use of Statistics), (2004)
3. Improving imputation method

3. 1 Current method & problem

【Current Method】

- Annual projections such as sales, fixed investment and current profits are aggregated and calculated into year-to-year % changes.
- Under the current imputation method, nonresponse is substituted using a value of the previous year from the same enterprise. Therefore, year-to-year % changes of annual projections of unanswered enterprises become 0%.

<table>
<thead>
<tr>
<th></th>
<th>FY2010</th>
<th>FY2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Investment</td>
<td>150</td>
<td>N/A</td>
</tr>
</tbody>
</table>

150 is substituted

【Problem】

- When the economy changes drastically, the changes may be underestimated (year-to-year % changes of annual projections come closer to ±0%).
- The results of the Tankan survey after the Lehman shock in 2008 did not appear to adequately reflect accurate business conditions.

→ The current imputation method should be reconsidered!
3. Improving imputation method

3. 2 alternatives

<table>
<thead>
<tr>
<th>Imputation method</th>
<th>Outline of the Imputed Value</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>① Previous Value Imputation</td>
<td>value of the previous year (or month) from the nonrespondent</td>
<td>Tankan Survey (March 2004 ~ )</td>
</tr>
<tr>
<td>② Growth Rate Imputation</td>
<td>value of the previous year (or month) from the nonrespondent multiplied by the growth rate derived from respondents</td>
<td>Current Survey of Commerce</td>
</tr>
<tr>
<td>③ Ratio Imputation</td>
<td>another item value of the nonrespondent multiplied by the ratio to another item derived from respondents</td>
<td>U.S. Public Libraries Survey</td>
</tr>
<tr>
<td>④ Mean Imputation</td>
<td>average value of respondents</td>
<td>Labour Force Survey</td>
</tr>
<tr>
<td>⑤ Regression Imputation</td>
<td>predicted value by using a regression model</td>
<td>U.K. Annual Business Survey</td>
</tr>
<tr>
<td>⑥ Donor Imputation</td>
<td>selected donor’s value in certain ways</td>
<td>Canadian Annual Survey of Manufactures</td>
</tr>
</tbody>
</table>

④～⑥: not appropriate for the Tankan survey

④ ・・・ Less accurate than the current method (i.e., ① Previous Value Imputation) according to the past research paper by the BoJ
⑤ ・・・ Heavy system load
⑥ ・・・ Samples of the Tankan survey vary in size.

①～③ methods are compared in the following simulation exercise.
3. Improving imputation method

3. 3 Comparing current method & alternatives

<Three Imputation Methods>

<table>
<thead>
<tr>
<th>Imputation Method</th>
<th>Imputed Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>①Previous Value Imputation</td>
<td>value of the previous year from the nonrespondent</td>
</tr>
<tr>
<td>②Growth Rate Imputation</td>
<td>value of the previous year from the nonrespondent × growth rate derived from respondents</td>
</tr>
<tr>
<td>③Ratio Imputation</td>
<td>another item value of the nonrespondent × ratio to another item derived from respondents</td>
</tr>
</tbody>
</table>

- Comparing the statistical accuracy between three imputation methods, we conduct Monte Carlo simulation.

  ① Creating “artificial” missing values to the original complete dataset of all sample enterprises & imputing these missing values in each method (500 times)

  ② Calculating RRMSE (Relative Root Mean Squared Error) of each method

  \[
  RRMSE = \frac{\sqrt{MSE}}{\theta} \quad MSE = E\{(\theta - \hat{\theta})^2\}
  \]

  Aggregates of the original complete dataset

  Aggregates of the dataset after imputing “artificial” missing values

  Mean Squared Error

  The difference between \( \theta \) & \( \hat{\theta} \)

  ③ Adopting the method with the smallest RRMSE as an alternative
### 3. Improving imputation method

#### 3. 3 Comparing current method & alternatives

<table>
<thead>
<tr>
<th>Survey Items</th>
<th>Imputed Value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>Current Method</td>
<td>Alternatives</td>
</tr>
<tr>
<td></td>
<td>①Previous Value Imputation</td>
<td>②Growth Rate Imputation</td>
</tr>
<tr>
<td></td>
<td>$S_i^{(t-1)}$ ⋅⋅⋅ Sales</td>
<td>$S_i^{(t-1)} \times \sum S_k^{(t)}$</td>
</tr>
<tr>
<td>Fixed Investment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$I_i^{(t-1)}$ ⋅⋅⋅ Investment</td>
<td>$I_i^{(t-1)} \times \sum I_k^{(t)}$</td>
</tr>
<tr>
<td></td>
<td>$P_i^{(t-1)}$ ⋅⋅⋅ Profits</td>
<td></td>
</tr>
<tr>
<td>Current Profits</td>
<td>③Ratio Imputation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$S_i^{(t)} \times \left( \frac{P_i^{(t-1)}}{S_i^{(t-1)}} + \left( \frac{\sum P_k^{(t)}}{\sum S_k^{(t)}} - \frac{\sum P_k^{(t-1)}}{\sum S_k^{(t-1)}} \right) \right)$</td>
<td></td>
</tr>
</tbody>
</table>

* i : nonrespondent,  k : respondent,  t : current year,  t-1 : previous year

※ The range in which the average growth rate or the average ratio to another item of respondents is calculated is defined by “size of the enterprise” × “branch of industry”.

*1 ③Ratio Imputation is passed over because there is not an appropriate ratio to another item in the case of imputation of Sales and Fixed Investment.

*2 ②Growth Rate Imputation is passed over because the growth rate cannot be defined when Current Profits become negative.
3. Improving imputation method

3. 4 Comparing RRMSEs

【Sales】

When the year-to-year % change showed a drastic movement such as FY2009, the RRMSEs of ①Previous Value Imputation are quite large.

In other years, the RRMSEs of ①Previous Value Imputation are larger than those of ②Growth Rate Imputation in general.

②Growth Rate Imputation is better.

Comparing RRMSEs

In other years, the RRMSEs of ①Previous Value Imputation are larger than those of ②Growth Rate Imputation in general.
3. Improving imputation method

3. 4 Comparing RRMSEs

【Fixed Investment】

When the year-to-year % change showed a drastic movement such as FY2009, the RRMSEs of ① Previous Value Imputation are quite large in some categories.

In other years, the RRMSEs of ① Previous Value Imputation are not so different from those of ② Growth Rate Imputation.

② Growth Rate Imputation is better.
3. Improving imputation method
3.4 Comparing RRMSEs

【Current Profits】

When the year-to-year % change showed a drastic movement such as FY2008, the RRMSEs of ① Previous Value Imputation are quite large.

In other years, the RRMSEs of ① Previous Value Imputation are not so different from those of ③ Ratio Imputation.

③ Ratio Imputation is better.
### 3. Improving imputation method

#### 3. 3 Comparing current method & alternatives

<table>
<thead>
<tr>
<th>Survey Items</th>
<th>Imputed Value</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sales</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Previous Value Imputation | Growth Rate Imputation | RRMSE
| $S_i^{(t-1)}$ | $S_i^{(t-1)} \times \frac{\sum S_k^{(t)}}{\sum S_k^{(t-1)}}$ | ①Previous Value Imputation |
| $I_i^{(t-1)}$ | $I_i^{(t-1)} \times \frac{\sum I_k^{(t)}}{\sum I_k^{(t-1)}}$ | $>$ |
| *1 Growth Rate Imputation | RRMSE | ②Growth Rate Imputation |
| **Fixed Investment** | | |
| | | |
| **Current Profits** | | |
| Ratio Imputation | | RRMSE
| $S_i^{(t)} \times \left( \frac{P_i^{(t-1)}}{S_i^{(t-1)}} + \left\{ \frac{\sum P_k^{(t)}}{\sum S_k^{(t)}} - \frac{\sum P_k^{(t-1)}}{\sum S_k^{(t-1)}} \right\} \right)$ | ③Ratio Imputation | *2 |
| | | |

* i : nonrespondent , k : respondent , t : current year, t-1 : previous year

* The range in which the average growth rate or the average ratio to another item of respondents is calculated is defined by “size of the enterprise” × “branch of industry”.

*1 ③Ratio Imputation is passed over because there is not an appropriate ratio to another item in the case of imputation of Sales and Fixed Investment.

*2 ②Growth Rate Imputation is passed over because the growth rate cannot be defined when Current Profits become negative.
4. Conclusion

➢ The current imputation method should be revised.
  — The current imputation method (i.e., Previous Value Imputation) appears to underestimate annual changes when the economy fluctuates drastically.
  — Based on our simulation exercise, the alternative imputation method is likely to improve the statistical accuracy.

➢ Remaining challenges:
  (1) Revision of survey items
      e.g., medium / long-term inflation expectations
  (2) Sample design
      Introduction of the Economic Census (scheduled in Feb. 2012)