Current Status & Future of Statistical Reform – Utilization of Real-Time Data & Alternative Data

By Nariyasu Yamasawa

Author
Nariyasu Yamasawa

Foreword

Enhanced statistics are important in the promotion of Evidence-Based Policy Making (EBPM). Fiscal 2022 is the fifth year of the period covered in the "Master Plan Concerning the Development of Official Statistics". In the meantime, an inadequate statistical process was discovered in the Monthly Labour Survey and the Construction Orders Survey, and mistrust of statistics has yet to be resolved. However, statistical reform is making steady but sure progress, and the "Economic Conditions Survey" was conceived as one outcome.

On the other hand, the spread of the Covid-19 pandemic has had a huge impact on society, and has completely changed the scenery of data analysis. This article describes the progress of the two efforts on real-time database and alternative data.

Creation of "Economic Conditions Survey"

In 2017, the "Statistical Reform Promotion Council" was established and fundamental statistical reform began. Statistical reform became concrete in the form of newly creating the Master Plan Concerning the Development of Official Statistics which is a five-year working plan. The period covered is between fiscal 2018 and fiscal 2022. Fiscal 2022 is the last year of the Basic Plan and also the year to formulate the plan for the next period (fiscal 2023 to fiscal 2027).

One of the achievements of this period was the creation of the Economic Conditions Survey. The highlight of the statistical reform is improving the accuracy of GDP statistics and the newly created statistics aim to realize this goal.

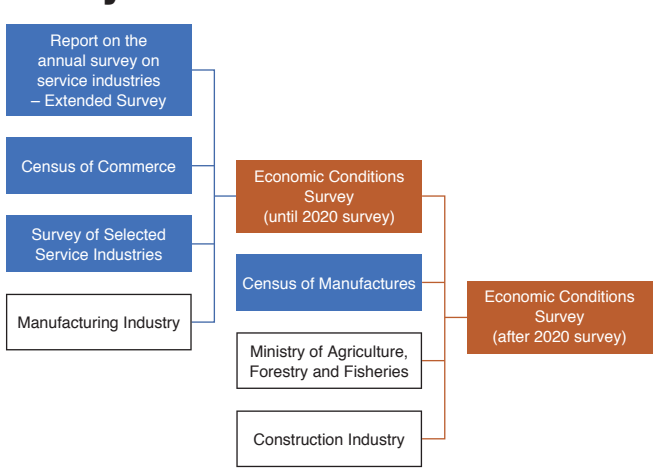
The three features of the statistics are: (1) speed of the cycle, (2) compatibility with GDP, and (3) survey of all industries. The first is the speed of the cycle at one year, which is fast for structure statistics. The most detailed statistics that comprehensively capture corporate activities are the "Economic Census for Business Activity". Census refers to a "complete survey" and it surveys all businesses. The statistics are accurate but the survey is only conducted once every five years. The gap between the years that are not surveyed (called the interim years) has been an issue. The Economic Conditions Survey will function as the statistics to fill these interim years.

The second is that the statistics are adequate in creating GDP. There is the "Financial Statements Statistics of Corporations by Industry" from the Ministry of Finance for corporate statistics, and figures such as sales, capital investment, and current profits can be grasped from these statistics. But in order to calculate GDP, more detailed information is required: information such as which industry used what products and how these products were supplied. The Economic Conditions Survey also surveys the input-output structure which indicates the inter-industry relationship.

The third is that the statistics cover all industries. Reflecting the vertically segmented bureaucracy, statistics have thus far been created separately for each industry. The service industry had the "Report on the Annual Survey on Service Industries" and the "Survey of Selected Service Industries", while the wholesale and retail industry had the "Census of Commerce". These three statistics were first consolidated, and then a survey for the manufacturing industry was also conducted to begin the Economic Conditions Survey *(Chart 1)*.

In 2022, the "Census of Manufacturers" will also be consolidated, surveys for the construction industry and agriculture will begin, and a survey of almost all industries is set to begin. With the introduction of the Economic Conditions Survey, data necessary for GDP for all industries will be available to be surveyed. Improvement in the accuracy of GDP statistics can be expected.

CHART 1
Transition of Economic Conditions Survey



Note: For 2021, "Economic Census for Business Activity" was carried out.
Source: Compiled by the author

Capturing Cost Structure & Non-Main Business Projects Now Possible

Let us look at what the Economic Conditions Survey can actually capture. Two surveys have been conducted so far: the 2019 survey (2018 was the scope of the survey) and the 2020 survey (2019 was the scope of the survey). Since there have only been two surveys, analysis by time-series is limited but detailed analysis of the economic structure is possible. Total value-added for all industries was 255.7 trillion yen. The highest percentage by major industry classifications was the manufacturing industry. This was followed by the wholesale industry and the retail industry. Next come industries that are around the same size. These are "health care and welfare services", "transportation industry and postal services", "telecommunications industry", "financial services and insurance business", and "academic research, and professional and technical services".

Cost structures are also surveyed and therefore features by industry can also be revealed. "Employment placement and worker dispatch business" and "medical services" are both thought to be labor-intensive industries, but there are differences in the breakdown. With "Employment placement and worker dispatch business", "total wages" were very high at 69.7% and "welfare benefits" for the employees take up 8.7%. It shows that personnel expenses alone account for close to 80%.

In looking at "medical services", on the other hand, "total wages" were relatively high at 45.6%, but the percentage is also high for pharmaceuticals and equipment with "medicine costs" at 12.3% and "costs of materials" at 10.2%.

In addition, the survey also looks at non-main business activities. Sales of Uniqlo as a retailer are large, but in terms of making garments in-house, it is also a manufacturer. In the Economic Conditions Survey, in addition to sales from its main business in retail, sales for each of the non-main businesses such as manufacturing are surveyed, and a snapshot which is closer to reality can be captured.

Real-Time Data & Alternative Data

Next, new developments brought on by the spread of novel coronavirus infections will be explained. These are the real-time data and alternative data.

First, let us look at real-time data. Real-time may be thought to be defined as "currently ongoing", but that is not what it means. Real-time data refers to data that is available at the time of making policy decisions. Statistics are revised whenever new information is obtained. When assessing past policies, analysis is oftentimes conducted using data that is available at the time of the analysis, but most times the data has been revised since the time the policy decision was made. When the spread of novel coronavirus infections was expanding, the number of newly infected patients, ways of regulation, and economic activities all changed every passing moment. What is important when assessing policies later on is not to bring out the "*posteriori* argument". Making decisions based on all available past incidents and making decisions relying only on available data under future uncertainties are different. In order to conduct an accurate assessment, it is important to prepare and utilize real-time data.

While real-time data is being prepared primarily by central banks in Europe and the United States, such movements cannot be observed in Japan. A research team with the Tokyo Foundation for Policy Research to which I belong has been constructing and offering real-time data.

Major Revision to GDP During the Lehman Shock

The revision status of GDP has been tracked by using the real-time database of the Tokyo Foundation for Policy Research. Preliminary data is used when making policy decisions. Later, when the extent of the revisions becomes large, the initial policy decision becomes something that was made based on inaccurate information.

Let us first look at the GDP statistics during the Lehman Shock. *Table 1* follows the trend in how the real GDP growth rate changed at each of the release dates. The more to the right of the chart, the more up-to-date the data is. The real GDP growth rate for the period July-September 2008 was initially an annual rate of negative 0.4% year-on-year for its First preliminary release, but a year later it was revised downwards to a negative 6.5%. When future analysts analyze the policy responses to the Lehman Shock, they may wonder "why policy responses were delayed when they knew of the sharp drop for the period July-September 2008". However, what was first revealed was the slight drop in GDP. Conducting policy assessment using revised data can lead to errors in judgement.

The First preliminary value for the period July-September 2009 was a very high growth of 4.8%, but a year later it was a negative 1.5%. This implies that the growth rate which was thought to be positive was actually negative, and preliminary figures can no longer be trusted. The size of the revision is one reason behind the mistrust

TABLE 1

GDP statistics during the Lehman Shock (%)

	as of	2008:3	2008:4	2009:1	2009:2	2009:3	2009:4	2010:1	2010:2	2010:3	2010:4
Real GDP growth rate (seasonally adjusted year-on-year growth rate)	2008/7-9.	-0.4	-2.3	-2.5	-3.9	-6.5	-4.3	-4.2	-5.4	-5.4	-5.0
	10-12.		-12.7	-14.4	-13.1	-11.5	-11.4	-9.6	-10.0	-10.4	-10.9
	2009/1-3.			-15.2	-11.7	-12.2	-12.3	-15.9	-16.6	-15.8	-20.1
	4-6.				3.7	2.7	5.2	7.4	10.4	9.9	10.8
	7-9.					4.8	0.0	0.5	-1.0	-1.5	-1.9
	10-12.						4.6	4.2	4.1	4.2	7.3
	2010/1-3.							4.9	4.4	6.6	6.0
	4-6.								0.4	1.8	2.1
	7-9.									3.9	3.3
	10-12.										-1.1

Note: The period as of was upon the release of the First preliminary results. For example, 2008:3 denotes figure as of the release of the First preliminary results for the period July to September 2008.

Source: The Tokyo Foundation for Policy Research real-time database

of GDP statistics.

Size of Revision Shrinks During Coronavirus Infections

During the period when the spread of novel coronavirus infections was expanding, major revisions were not conducted for a year after the First preliminary results were released. The deviation between the First preliminary results and the Second preliminary results especially have been kept small compared to the period of the Lehman Shock. With the spread of novel coronavirus infections expanding, the April-June 2020 data, when a State of Emergency was declared for the first time, was initially released to be a decline by an annual rate of 27.8% year-on-year, but the rate of decline remained approximately the same after 18 months *(Table 2)*.

There are two reasons for this. The first is the ingenuity of the seasonal adjustment method. GDP statistics include increases and decreases by season, and therefore there is a need to create seasonally adjusted values which exclude this element. The reason for seasonally adjusted values greatly swinging during the Lehman Shock was because the large drop due to the turmoil of the financial market was processed as seasonal fluctuation. Past values are revised to cancel out seasonal fluctuations. This time, the big shock was processed as a temporary decline not through seasonal fluctuations, and therefore there were no major revisions to seasonally adjusted values. This is an example where improvement proved successful.

The other is that the First preliminary results utilized various data. Upon the release of the First preliminary results, some data did not have enough information for the entire quarter (three months). This is when there is two months' worth of data but data for the third month is not yet public. Originally, the growth rate for the two months has been applied to estimate the three months of data. Then there came cases such as the period January-March 2020, when January and February were not affected by the spread of novel coronavirus infections but March was greatly affected. If a similar numerical value as January and February were to be used for March, it would deviate greatly from reality.

Thus, estimations were devised by using industry data and data that was available only from large corporations. For example, the air transport industry has official statistics called the "Annual Report on Air Transport Statistics", but data collected by interviewing the major airline companies was used. For the food service industry, data collected by the industry organization was utilized. By doing so, the size of the revision between the First preliminary and the Second preliminary results remained small. January-March 2020 data was initially a negative growth of 3.4%, but most recently it has been revised to 1.8% growth, and while small, the positive/ negative sign has changed to be positive, and there is a need to further verify how much it will be revised.

Utilization of Alternative Data

What is anticipated for utilization is alternative data. Alternative data is substitute data which is not official statistics. It is like utilizing information generated from businesses as byproducts, so to speak, as they conduct economic activities.

What first came to be utilized with the spread of novel coronavirus infections was data on the workforce. When infections expand, there is a need to restrict economic activities depending on the status of the infections, and there is also a need to find out how much of the economy was suppressed by that restriction. Statistics such as GDP which is released at a delayed timing and can only be grasped on a quarterly basis are not enough to respond to that.

In relation to government statistics, the Ministry of Economy, Trade and Industry has constructed a system which allows results of the Census of Commerce, official statistics, to be known early by using the POS (Point of Sale information management system) data, and it is called the "METI POS – Indices for retail sales amount (micro indices)".

Weekly sales statistical data can be gained and since data can also be obtained by region and business categories (convenience stores, mass home electronics retailers, supermarkets, and others), it is important data to find the most up-to-date trends in consumption. Looking at sales by business categories, supermarket sales grew with the stay-at-home demand immediately after the novel coronavirus infections expanded, but by 2021 the growth has been stable also as a reaction. Sales for convenience stores were sluggish with people not going in to offices when working from home, but as workforces recovered, growth gradually turned positive. Sales for mass home electronics retailers and home-centers have recently been mostly declining year-on-year, but in contrast sales for drug stores have been increasing.

V-RESAS also Revolutionary

V-RESAS, which was constructed after the spread of coronavirus

TABLE 2

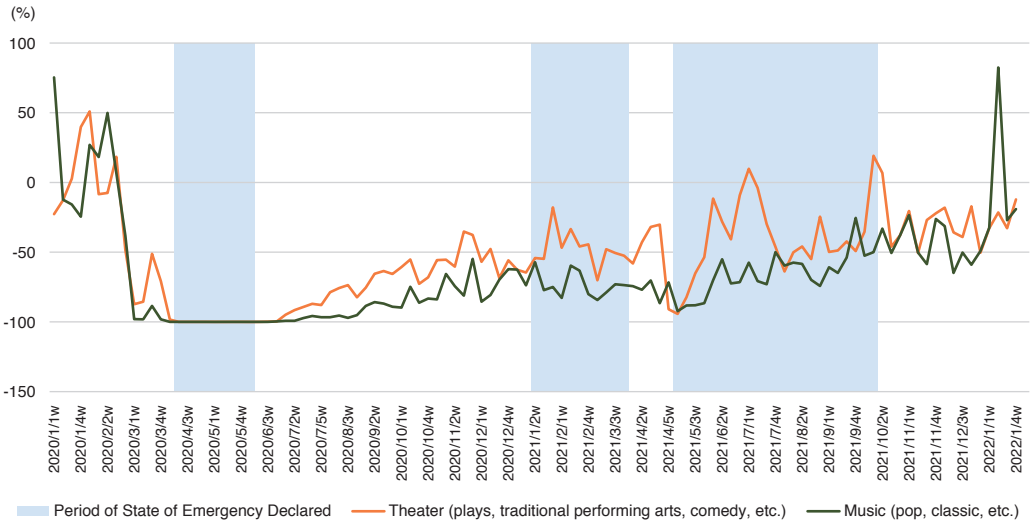
Minor revisions during the spread of Covid-19 (%)

	as of	2019:3	2019:4	2020:1	2020:2	2020:3	2020:4	2021:1	2021:2	2021:3	2021:4
Real GDP growth rate (seasonally adjusted year-on-year growth rate)	2019:7-9	0.2	0.5	0.0	0.2	0.2	0.7	0.5	0.5	0.5	0.1
	2019:10-12		-6.3	-7.3	-7.0	-7.1	-7.1	-7.4	-7.5	-7.6	-10.6
	2020:1-3			-3.4	-2.5	-2.3	-2.2	-1.9	-2.3	-2.3	1.8
	2020:4-6				-27.8	-28.8	-29.3	-28.6	-28.2	-28.2	-28.2
	2020:7-9					21.4	22.7	22.9	22.8	23.5	23.0
	2020:10-12						12.7	11.6	11.9	11.8	7.5
	2021:1-3							-5.1	-3.7	-4.1	-2.1
	2021:4-6								1.3	1.5	2.4
	2021:7-9									-3.0	-2.7
	2021:10-12										5.4

Note: The period as of was upon the release of the First preliminary results. For example, 2019:3 denotes figure as of the release of the First preliminary results for the period July to September 2019.
Source: The Tokyo Foundation for Policy Research real-time database

CHART 2

Recovery trend in ticket sales for various events



Note: Weekly, Growth from the 2019 same week in ticket sales. Created by using V-RESAS.
Source: PIA Corporation

infections, is also a revolutionary data analysis system. It is offered by the Cabinet Secretariat Council for the Realization of the Vision for a Digital Garden City Nation Secretariat and the Cabinet Office Promotion Office for Vitalizing Local Economy. It is a website created by collecting various alternative data.

RESAS is a system which analyzes regional economies such as prefectures and municipalities, and V-RESAS was created as a version to respond to the novel coronavirus. "V" in V-RESAS stands for vital signs – a generic term used for pulse and body temperature. It signifies frequently used data which reflects economic activities. Fast data can be obtained in addition to allowing data view by regions.

For the turnout of people which impacts the pace of the spread of infections, GPS information from smartphones was used to capture populations that stayed at various locations. For consumption trends, there is the credit card payment information used as data for those on the pay-side, and there is the POS (Point of Sale) data at supermarkets as data for the sell-side.

Declarations of a State of Emergency and Quasi-State of Emergency have had a grave impact on the activities of restaurants. The situation of their activities can be estimated by using the view count of restaurant information websites. In addition to the entire sector, individual activities can be viewed by different genres such as "Western cuisine", "Cafés and sweets", "Family restaurants and fast food", and "*Izakaya* and bars". The impact on the tourism industry is also large, and trends in the number of lodgers can also be viewed.

In relation to the employment environment, the number of job offers that were published can be tracked by industry classifications such as construction, manufacturing, and distribution, and moreover, information can also be grasped by job types such as specialist jobs and sales and services.

Chart 2 is the status of ticket sales which shows trends in events. April-May 2020 when the State of Emergency was declared throughout Japan saw almost all activities coming to a halt, and shows a negative 100%. Later, activities slowly began to expand with minimal regard to the State of Emergency declaration. There were some weeks that temporarily exceeded the level of 2019 which was before the spread of coronavirus infections.

These data do not cover all activities, and therefore their reliability falls short of that of official statistics. But they allow us to catch signs of economic activities before accurate data is released, and in this sense can be said to be useful data.

Conclusion

This article has described how GDP statistics were improved to promote EBPM. The big achievement was the establishment of the Economic Conditions Survey. It will become the statistics at the core of GDP estimations in the future.

Since the spread of coronavirus infections, enhancement of real-time data and alternative data became a challenge. With the use of real-time data, data that was available at the time when policy decisions were made can be identified and should be useful in conducting policy assessments. Especially during the period when the spread of coronavirus infections was fast expanding, the situation was changing rapidly and thus it is important to record data that was used at the time. Similarly, during the period when the spread of coronavirus infections was fast expanding, the importance of alternative data heightened. Going forward, using alternative data to complement official statistics will be useful in analyzing the current situation. JS

Nariyasu Yamasawa is a professor of the Faculty of Management at Atomi University. He has worked for the *Nihon Keizai Shimbun (Nikkei)* and the Japan Center for Economic Research, and is a former director of the Statistics Commission Office at the Ministry of Internal Affairs and Communications.