Bank of Japan Working Paper Series

Development of "Alternative Data Consumption Index": Nowcasting Private Consumption Using Alternative Data

Tomohiro Okubo*
tomohiro.ookubo@boj.or.jp

Koji Takahashi**
kouji.takahashi-2@boj.or.jp

Haruhiko Inatsugu*
haruhiko.inatsugu@boj.or.jp

Masato Takahashi*
masato.takahashi@boj.or.jp

No.22-E-8
July 2022

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

* Research and Statistics Department
** Research and Statistics Department (currently at the Personnel and Corporate Affairs Department)

Development of "Alternative Data Consumption Index": Nowcasting Private Consumption Using Alternative Data*

Tomohiro Okubo† Koji Takahashi‡ Haruhiko Inatsugu§ Masato Takahashi#

July 2022

Abstract

In the field of macroeconomic analysis, there has recently been a growing interest in "alternative data" or nontraditional data whose information sources differ from those of existing statistics. Using alternative data that become timely available, this paper aims to capture developments in Japan's private consumption at the macro level earlier than existing statistics. We construct the "Alternative Data Consumption Index" (ALC) by combining three types of alternative data: (1) credit card transaction data (*JCB Consumption NOW*); (2) point-of-sale (POS) data (METI POS and GfK); and (3) spending records obtained from a personal financial management service (Money Forward). We nowcast the Consumption Activity Index (CAI), which is compiled and released by the Bank of Japan, using the ALC. With respect to timeliness, the ALC has a significant advantage over the CAI; the ALC for the month is available in the middle of the following month, approximately 3 weeks earlier than the release of the CAI. Our findings show that the ALC is generally accurate in nowcasting the CAI and thus aggregate consumption developments. It also accurately captures the substantial changes in consumption activities caused by the spread of COVID-19 since spring 2020. Overall, the results suggest that alternative data can capture macro level consumption activity promptly and accurately, making them a powerful tool for understanding economic conditions.

JEL classification: C49, E21, E27

Keywords: Nowcasting, Alternative Data, Private Consumption

---

* The authors thank Seisaku Kameda, Takuji Kawamoto, Tomohiro Sugo, Teppei Nagano, Koji Nakamura, Tomoyuki Yagi, and staff members of the Bank of Japan for their valuable comments. Any remaining errors are our own. We also thank Yuko Kobayashi for her assistance in preparing this English version. Money Forward, Inc. provided the de-identified, aggregated dataset of Money Forward ME transactions to the authors only for this research purpose. The views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan or Money Forward, Inc.

† Research and Statistics Department, Bank of Japan. E-mail: tomohiro.ookubo@boj.or.jp

‡ Research and Statistics Department (currently at the Personnel and Corporate Affairs Department), Bank of Japan. E-mail: kouji.takahashi-2@boj.or.jp

§ Research and Statistics Department, Bank of Japan. E-mail: haruhiko.inatsugu@boj.or.jp

Research and Statistics Department, Bank of Japan. E-mail: masato.takahashi@boj.or.jp

1. Introduction

There has recently been a growing interest in "alternative data" or nontraditional data whose information sources differ from those of existing statistics in the field of macroeconomic analysis. Alternative data are massive amounts of data collected and processed by advanced information technology based on various socioeconomic activities, such as those of individuals or firms. These massive amounts of data include (1) consumption data based on credit card transactions, (2) mobility data compiled using location information collected from smartphones, and (3) various types of textual data.

Many of these alternative data can be compiled in a short period compared with the existing statistics, which require certain steps, such as data collection and aggregation before they are released. Alternative data sources have been used to assess macroeconomic conditions, reflecting the need to capture the current state of the economy as promptly as possible. Such trends have accelerated in the current environment, where economic activities are changing rapidly due to the impact of COVID-19. Hence, policymakers, for example, have been conducting active research on "nowcasting" economic conditions using alternative data.

One of the areas where efforts have been made toward nowcasting using alternative data is private consumption, which has been most severely affected by the COVID-19 spread. During the implementation and subsequent lifting of public health measures, private consumption has fluctuated significantly at a historical rate that the existing monthly or quarterly statistics could not keep up with. In such cases, efforts have been made to capture consumption activities earlier and more accurately than existing statistics by using alternative data, such as credit card transaction data or location information collected from smartphones, which can be updated daily or weekly.

For example, Dunn *et al.* (2020, 2021) stated that daily credit card transaction data are useful for forecasting the *Monthly Retail Trade Survey* in the United States. Alternative data are also deemed useful in many other countries, as evidenced by research indicating that card transaction data help improve the accuracy of nowcasting consumption (e.g., Chapman and Desai [2021, Canada]; Bounie *et al.* [2020, France]; Carvalho *et al.* [2021, Spain]).

Japan has also made progress in nowcasting private consumption using alternative data. For example, after confirming their high correlation with consumption trends for face-to-face services, the Bank of Japan (BOJ) has been assessing current consumption trends using daily mobility data (e.g., the nighttime population of selected downtown areas in Tokyo, mobility trends for places such as restaurants, shopping centers, and theme parks), which some private companies have begun to release since the outbreak. Furthermore, the BOJ has been using alternative data based on actual expenditure to promptly grasp consumption activities in Japan; specifically, it uses *JCB Consumption NOW*, an index of credit card transaction data, and METI POS, the aggregate data of point-of-sale (POS) purchases of goods or services. Such analysis results were published in the *Outlook for Economic Activity and Prices* (Outlook Report), and some of the data have remained being published in the Outlook Reports as indicators to be periodically monitored for the time being.[1] Moreover, Matsumura *et al.* (2021) reported that data on the hourly population surrounding commercial and public facilities -- based on smartphone location information -- are useful in capturing consumption activities in the services industry (e.g., amusement parks, shopping centers, and the food services industry). Additionally, Urasawa (2022) indicated that *JCB Consumption NOW* can be used to forecast service consumption in real-time.

As these examples show, alternative data perform well in nowcasting private consumption. However, the definitions and data coverage of alternative data differ from those of existing statistics, and these differences should be considered when using alternative data. For example, credit card payment data cannot capture the consumption activities of individuals who do not use the cards covered by the data. Moreover, in Japan, credit cards are frequently used to pay for relatively expensive goods or services, whereas they are not considered suitable for daily small transactions or automobile purchases, for which loans are frequently obtained. Although POS data cover a wide range of purchases regardless of payment method, they are difficult to use for service consumption analysis. In addition, data on goods consumption in some physical stores are not included in POS data. Data obtained from personal financial management services have limitations as well, because the demographic distribution of users may be skewed toward individuals with a high level of digital literacy (mainly younger generation). Thus, alternative data have both advantages and disadvantages. Although they can accurately capture developments in specific industries or items, they are not necessarily functional for others, due to the difference in their information sources. Therefore, using a single set of alternative data is insufficient for comprehensively capturing macroeconomic developments in private consumption. To address this issue, employing multiple sets of alternative data in a mutually complementary manner is effective, maximizing the benefits of each dataset.

[1] The BOJ has been using alternative data for economic analysis prior to the outbreak of COVID-19. Kameda (2022) outlined the BOJ's analyses that employ alternative data, including those conducted after the outbreak. The BOJ also launched a new page on its website called "Alternative Data Analysis," where related research is posted.

By combining multiple sets of alternative data, this paper aims to accurately nowcast developments in private consumption in Japan. Alternative data have been used to nowcast private consumption at the item or industry level (e.g., face-to-face services industry); however, to the best of our knowledge, our study is the first to nowcast developments in consumption at the macro level using an index that combines multiple alternative datasets from various sources.[2] This paper intends to nowcast the Consumption Activity Index (CAI), which is compiled and released by the BOJ on a monthly basis to track developments in private consumption at the macro level.[3]

For nowcasting, we employ three types of alternative data: (1) a consumption index based on credit card transaction data (*JCB Consumption NOW*); (2) POS data (METI POS and GfK); and (3) spending records obtained from a personal financial management service (Money Forward, aggregate amount broken down by individual attribute). These data are obtained in approximately 2 weeks (e.g., data on consumption developments for April are available in mid-May), whereas existing statistics are released with at least a 1-month lag. Thus, alternative data outperform existing statistics with respect to timeliness. Moreover, unlike mobility data or various consumer confidence indicators, these alternative data are compiled based on actual individual expenditure or store sales and thus are quite useful for accurately capturing developments in private consumption. This study proposes the "Alternative Data Consumption Index" (ALC) by combining selected alternative data that best fit each item-level index in the CAI. The biases inherent in each set of alternative data are corrected as much as possible during the ALC construction process.

Our findings show that alternative data perform well in capturing developments in private consumption across many industries and items. Furthermore, the ALC, which is built by combining alternative data that have the best fit to each item in the CAI, is generally functional for estimating the CAI and thus aggregate consumption trends. It also accurately captures the substantial changes in consumption activities that have occurred as a result of the COVID-19 spread since spring 2020, prior to the release of the existing statistics.

[2] Although the focus of this paper is on nowcasting private consumption, some analyses are aimed at nowcasting GDP using alternative data. For example, noting that the BOJ's nowcasting models constructed using monthly data (primarily official statistics) for estimating Japan's GDP (Chikamatsu *et al.* 2021) are becoming less accurate during the pandemic, Nakazawa (2022) claimed that the models' accuracy improves when daily or weekly alternative data, such as Google Trends data and METI POS, are incorporated.

[3] For details on the CAI, see, for example, Nakamura *et al.* (2016) and Takahashi *et al.* (2021).

The remainder of this paper is organized as follows. Section 2 describes the alternative data used in our analysis. Section 3 provides an overview of the ALC calculation, and Section 4 discusses the ALC's overall performance. Finally, Section 5 concludes the paper.

2. Alternative data

This section outlines the three types of alternative data used in our analysis.

2.1. Credit card transaction data -- JCB Consumption NOW

The first set of alternative data is the credit card transaction data. We use *JCB Consumption NOW*, which is an index of JCB cards' transaction data, compiled and released by Nowcast Inc. It is a consumption index based on randomly selected JCB cardholder transaction data. The data cover approximately 10 million holders of JCB cards (limited to those issued directly by JCB), allowing consumption developments to be tracked by detailed category, such as gender, age, or region. When compiling item-level indices, Nowcast Inc. uses the *Population Estimates* to correct sample biases. It also publishes aggregate indices ("Total," "Retail," and "Service") by combining item-level indices. In doing so, they take into account the industry weights in the *Current Survey of Commerce* and the CAI, allowing the aggregate indices to more accurately capture consumption trends at the macro level. Furthermore, because credit cards are commonly used for online purchases, they are suitable for capturing online consumption. Existing statistics cannot adequately track online consumption, which is another advantage of *JCB Consumption NOW*.

2.2. POS data -- METI POS and GfK

Second, we employ "POS data," also known as scanner data. It includes information such as point of sale, prices, and sales volumes. *JCB Consumption NOW* data are demand-side (consumer-side), whereas POS data are supply-side (sales-side).

For the analysis, we use two types of POS data: (1) "METI POS," which is compiled by the Ministry of Economy, Trade and Industry, and (2) an index provided by GfK Japan (hereafter simply referred to as "GfK"). METI POS displays aggregate retail sales for five business types (supermarkets, home centers, large electronics and appliance stores, convenience stores, and drugstores) on a weekly and monthly basis, broken down by item (e.g., foods, general merchandise, and household electrical appliances) using POS data. We use monthly METI POS data as a candidate for capturing changes in "food and beverages" and "drugs, cosmetics, etc." in the CAI. Meanwhile, for the second type, although GfK only covers consumption of household electrical appliances, it follows that of a broader range of goods when compared to the corresponding data in the METI POS. Therefore, we use GfK as a candidate for tracking "household electrical appliances" in the CAI.

2.3. Spending records based on personal financial management services -- Money Forward

The third type of alternative data is the spending records obtained from a personal financial management service. For our analysis, we use payment records from users of "Money Forward ME," a personal financial management service provided by Money Forward, Inc. This dataset contains the aggregate amount of payments recorded by users, broken down by a wide range of items. Thus, the individuals cannot be identified. It falls under the category of demand-side data. Users who have their individual attributes identified or who use the service on a monthly basis will be covered. Although we describe the sample selection rules in details later, note that even after this sample selection process, approximately 310,000 users satisfy the conditions, far exceeding the sample size of the *Family Income and Expenditure Survey* (approximately 9,000).

It should be noted that, because the users of the services thus far have been skewed toward younger generations, simply aggregating their payment records may result in a trend that differs from the actual consumption trend in the economy. To address this issue, we use the following advantage of the dataset in our analysis. That is, our dataset contains the average amount of spending by item for various groups of users stratified by age and family structure, calculated using detailed and anonymized records of all users' spending.[4] Thus, we can aggregate the average amounts of spending for different groups for each item using Japan's true population distribution as weights, thereby improving the estimation accuracy of each item-level aggregate data.

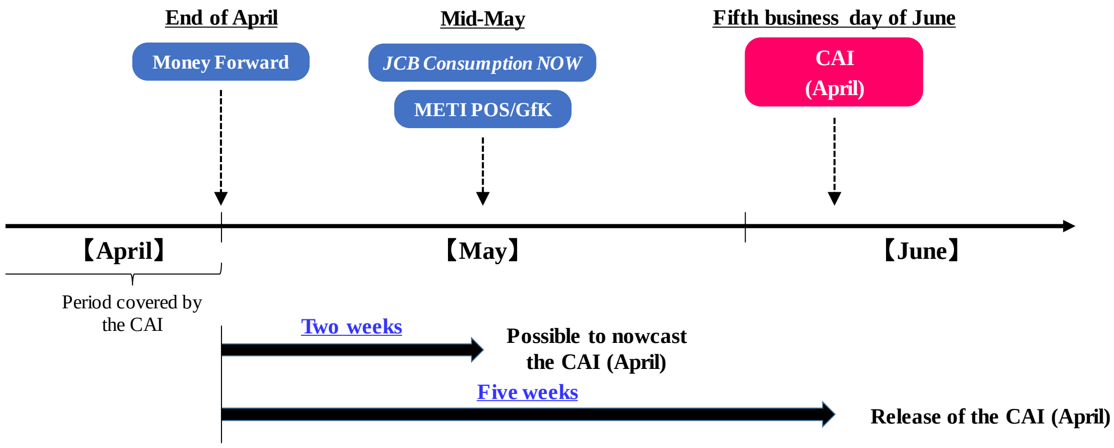
2.4. Timeliness and caveats of alternative data

All these three types of alternative data perform well with respect to timeliness. For example, alternative data for April consumption become available in mid-May. Meanwhile, the CAI for April, which this paper aims to nowcast, is released in early June because it takes time for its source statistics for April to be released and then aggregated (Chart 1). By using alternative data, we can reduce the time it takes to grasp developments in private consumption at the macro level, from approximately 5 to 2 weeks. Therefore, such data have an advantage in assessing economic conditions.

[4] The authors only set the thresholds, asking Money Forward, Inc. to aggregate the data.

Chart 1. Timeliness of Alternative Data

The Case of Nowcasting the CAI for April



However, these three types of alternative data exhibit dissimilar patterns of movement because their information sources differ. As previously stated, *JCB Consumption NOW* is based on credit card transaction data and thus accurately captures consumption of items that are more likely to be purchased using such cards than other payment methods. However, it cannot necessarily capture consumption in industries where credit card use is less frequent. The items covered by POS data are limited. Moreover, personal financial management service data cannot capture consumption of people who do not use the services. Furthermore, whether the data cover inbound tourism consumption (consumption in Japan by foreign visitors) or outbound tourism consumption (consumption overseas by Japanese residents) may influence movements in these alternative data (Chart 2). As will be discussed later in this paper, such differences in data coverage appear to have an impact when capturing developments in consumption since the outbreak of COVID-19, particularly for travel and accommodation services.

Noteworthily, alternative data may have some biases. For example, based on credit card transaction data or spending records obtained from personal financial management services, the aggregate amount of consumption may become higher or lower than the actual amount. This may occur if (1) a broader range of goods are purchased using credit cards amid the increased trend in cashless payments, or (2) consumption activity of people who use personal financial management services differs significantly from that of those who do not use such services -- for example, users of such services may be eager to restrain their spending.[5]

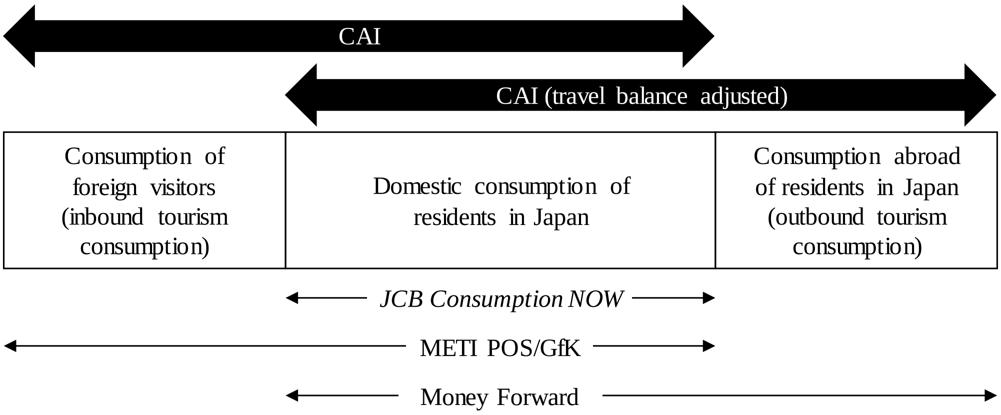
---

[5] If the rate of increase in credit card usage is moderate, the resulting biases will be less significant and thus more easily corrected. However, when online consumption expands rapidly and discontinuously, as seen in the current phase of COVID-19 spread, the biases caused by increased use of credit cards as the primary payment method for online purchases become more significant than those seen under normal

Therefore, when capturing developments in macro level private consumption using alternative data, the biases inherent in such data must be considered.

As previously discussed, developments in macro level private consumption are difficult to track when only one type of alternative data is used. This is because each of these three types of alternative data has its advantages in capturing consumption trends in specific industries or items and contains its own biases. To capture changes in macro level consumption as accurately as possible, it is desirable for researchers to combine multiple sets of alternative data while paying attention to their biases.

Chart 2. Coverage of the Alternative Data Used in the Analysis

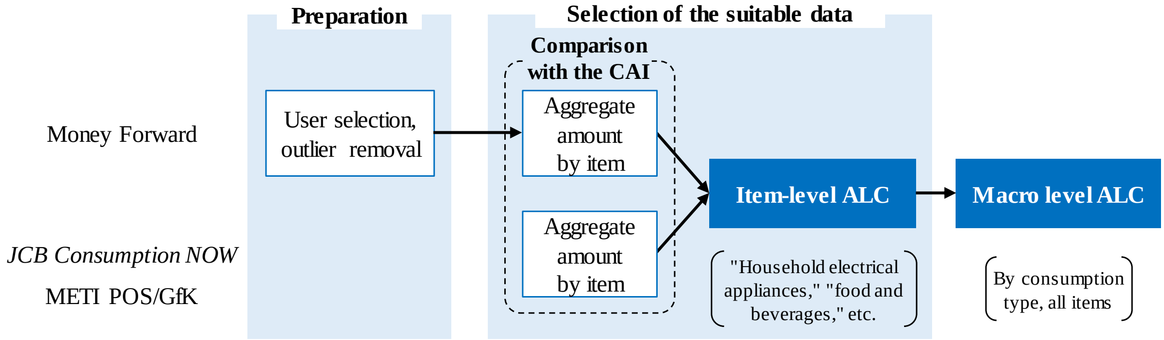


3. Calculation of the ALC

This section explains the methodology for calculating the ALC. We construct the ALC by the following steps (Chart 3). First, we aggregate the Money Forward data, which are broken down by individual attribute, in a manner consistent with Japan's population distribution, thus creating the "Money Forward Index" (MFI) for each item. Then, for each item, we choose the alternative data that best fit the corresponding series in the CAI (i.e., the item-level ALC). Finally, we combine the item-level ALC with the CAI weights to calculate the macro level ALC.

conditions and should thus be corrected with caution. For developments in online consumption since the outbreak of COVID-19, see Nakajima *et al.* (2022).

Chart 3. Construction of the ALC



3.1. Calculation of the MFI

As previously stated, the Money Forward data are divided into groups based on individual characteristics, such as age or gender. These data should be aggregated consistently with the distribution of the Japanese population before calculating the ALC.

First, we narrow the users down. Money Forward users include those who (1) use the service infrequently (update their spending records only occasionally), (2) have data that deviate significantly from the average (with exceptionally high income and spending), and (3) appear to be business-related. These users are excluded from the sample because they may introduce noise into the analysis of macro level private consumption developments. Moreover, to take account of the Japanese population distribution, we only collect data from users whose age and family structure are known. We specifically limit users based on the following seven conditions (Chart 4). Despite such stringent requirements, the total number of selected users turns out to be approximately 310,000.

Chart 4. Calculation of the MFI: Conditions for Selecting the Users

1. Two years or more have passed since the account was created
2. Uses the service every month
3. Connects more than two bank accounts
4. Does not use a corporate bank account
5. Has not used the services for businesses provided by Money Forward, Inc.
6. 10 million yen or less of spending and income per month, respectively
7. Individual attributes are identified

The total amount spent on each item is then calculated using the data of the users who satisfy the criteria. Specifically, after aggregating these data by user age and family structure, we compute a weighted average based on Japanese population distribution. Let g be a group of users of particular age and family construction, $E_{g,i,t}$ be the simple aggregate of spending for item i by this group at period t, and POP_g be the Japanese population of the same age and family structure as those of the users in group g (based on the *Population Census* for CY 2015). The amount of spending for item i at period t can be defined as follows:

$$\mathrm{MFI}_{i,t} = \sum_{g \in G} w_g E_{g,i,t} \quad \left(w_g := \frac{\mathrm{POP}_g}{\sum_{g' \in G} \mathrm{POP}_{g'}} \right). \tag{1}$$

The sigma notation in Equation (1) indicates that all groups' spending is aggregated. This is the total amount spent on item i at period t, as suggested by the Money Forward data. This total amount is referred to as the MFI for this item.

On the right-hand side of Equation (1), $E_{g,i,t}$, which denotes the amount of spending of each group, frequently fluctuates, resulting in a significantly biased aggregate amount. Such fluctuations may occur, for example, when some users make extremely large payments. These fluctuations are excluded from the calculation because they are considered outliers that deviate from developments in private consumption at the macro level. The effects of such outliers are specifically removed as follows. The contribution of changes in spending of group g, denoted as $e_{g,i,t}$, to the year-on-year rate of change in the MFI for item i, denoted as $\left(\mathrm{MFI}_{i,t} - \mathrm{MFI}_{i,t-12}\right)/\mathrm{MFI}_{i,t-12}$, is given as:

$$e_{g,i,t} := \frac{w_g \left(E_{g,i,t} - E_{g,i,t-12}\right)}{\mathrm{MFI}_{i,t-12}}. \tag{2}$$

If the contribution $e_{g,i,t}$ is in the upper or lower one percentile of the distribution $\{e_{g,i,t}\colon\ g \in G\}$, this group is excluded from calculation. We recalculate Equation (1) after excluding such groups because the population weights should be updated in accordance with changes in the number of groups included in the calculation. The ALC is then built using the MFI, from which the effects of outliers are removed.

3.2. Construction of the ALC

This section examines the fit between the three indicators -- *JCB Consumption NOW*, METI POS/GfK, and the MFI -- and the CAI items. Moreover, we choose the alternative data that best fit each item.[6]

---

[6] Conducting a multiple regression analysis using all three types of alternative data can be an option; however, if their movements are similar, a multicollinearity problem may arise, so we use a method based on a single regression analysis.

Using a single regression analysis, we assess the fit between these alternative data and the CAI items. The year-on-year rate of change in each of the three indicators is regressed to that in the CAI (nominal basis, not seasonally adjusted), as shown in Equation (3). For each item, the indicator (alternative data) with the highest R-squared with the CAI is selected. Each of these selected indicators is referred to as the item-level ALC; note that the ALC is based on year-on-year rates of change rather than levels. The sample period for the single regression analysis ranges from January 2018 to the most recent data.[7]

$$\left(\frac{\mathrm{CAI}_{i,t}}{\mathrm{CAI}_{i,t-12}} - 1\right) \times 100 = \alpha_i + \beta_i X_{i,t} + \varepsilon_{i,t} \tag{3}$$

$$\left(\begin{array}{c} X_{i,t}\text{: (1) } \textit{JCB Consumption NOW}\text{, (2) METI POS/GfK, or (3) MFI.} \\ \text{[year-on-year rate of change (\%)]} \end{array}\right)$$

In the case of automobiles, we use the number of new passenger car registrations, that is, the source statistics of the CAI, for the ALC because it is available earlier than the alternative data.[8] There are also some cases where R-squared for specific items, such as "life insurance" and "public broadcasting," with the CAI results to around zero, or the regression coefficient β_i between these items and the CAI is negative. This paper does not discuss whether the source statistics or alternative data are more accurate, or the reasons for such deviations, because that is not our primary goal. However, using alternative data for items that do not fit well with such data is not appropriate to forecast the CAI as accurately as possible. Instead of alternative data, the values obtained by mechanically extrapolating the CAI source statistics using the seasonal ARIMA (SARIMA) are employed for the ALC. In doing so, the SARIMA model's order is selected using the Akaike information criterion (AIC). The data available for selection range from January 2003 to the most recent data.

We can obtain the item-level ALC using the aforementioned procedure. Even if the year-on-year rates of change in the ALC and CAI for each item show similar trends, their averages or standard deviations do not necessarily coincide. To nowcast the CAI as accurate as possible, such differences should be corrected. Thus, we adjust the ALC as follows using coefficients obtained from a single regression in Equation (3).

[7] The best alternative data for each item are nearly identical to those from the pre-pandemic sample period (from January 2018 to March 2020).

[8] Alternative data can nowcast "automobiles" in the CAI, but these data cannot accurately capture the number of new passenger car registrations. This is because, for example, people do not typically use credit cards when purchasing automobiles, and thus *JCB Consumption NOW* is not suitable for this item. Money Forward data are also ineffective because they include components not covered by the source statistics of the CAI, such as used car purchases or automobile loan payments.

$$\mathrm{ALC}_{i,t}^{\mathrm{adjusted}} = \hat{\alpha}_i + \hat{\beta}_i \times \mathrm{ALC}_{i,t} \tag{4}$$

In Equation (4), $\hat{\alpha}_i$ and $\hat{\beta}_i$ represent the coefficients obtained from a single regression in Equation (3). Hereafter, the ALC refers to the one that is adjusted by Equation (4).

Finally, we combine the item-level ALC to create the ALC for each consumption type (durable goods, non-durable goods, services, and all items). This is the macro level ALC. When combining the ALC, we compute the level of the ALC for each item so that we can use the same calculation method as for the CAI. We go through the following steps in particular.

The ALC is based on year-on-year rates of change; therefore, we use Equation (5) to compute its level and thus nowcast the level of the CAI for each item.

$$\mathrm{CAI}_{i,t}^{\mathrm{nowcast}} = \mathrm{CAI}_{i,t-12} \times \left(1 + \frac{\mathrm{ALC}_{i,t}^{\mathrm{adjusted}}}{100}\right) \tag{5}$$

The level of the predicted value (index) of each of the three consumption types (durable goods, non-durable goods, and services) is calculated as the weighted average of the level of the predicted value of the CAI for each item ($\mathrm{CAI}_{i,t}^{\mathrm{nowcast}}$), which is obtained in Equation (5), using the item weights in the CAI. The sigma notation in Equation (6) indicates that indices for all items that comprise each consumption type are aggregated using item i's share in the consumption type to which it belongs (based on the CAI) as weights. The share of item i is denoted as w_i^{CAI}. The ALC for this consumption type is defined as the year-on-year rate of change in this index (the rate of change from the "genuine" CAI, which is denoted as ALC_t in Equation (7)). The ALC for all items is calculated by the year-on-year rate of change in the weighted average of the levels of these three consumption indices, using weights from the National Accounts of Japan.

$$\mathrm{CAI}_t^{\mathrm{nowcast}} = \sum_i w_i^{\mathrm{CAI}} \mathrm{CAI}_{i,t}^{\mathrm{nowcast}} \tag{6}$$

$$\mathrm{ALC}_t = \left(\frac{\mathrm{CAI}_t^{\mathrm{nowcast}}}{\mathrm{CAI}_{t-12}} - 1\right) \times 100 \tag{7}$$

The ALC can be calculated using the methods described earlier. Chart 5 depicts the CAI's source statistics. We intend to nowcast the items with CAI total ratios greater than 70% using only three types of alternative data.[9]

[9] The following items were excluded from our nowcasting analysis using alternative data: life insurance

Chart 5. Items Covered for Nowcasting Using Alternative Data

Consumption type	Item	Source statistics	Item weights
Durable goods	Household electrical appliances	Current Survey of Commerce	4.73
Non-durable goods	Food and beverages	Current Survey of Commerce	19.59
	Drugs, cosmetics, etc.	Current Survey of Commerce	6.50
	Clothes	Current Survey of Commerce	4.85
	Electricity, gas, and water	Survey of Electric Power Statistics, etc.	4.03
	Fuel	Current Survey of Commerce	2.75
	Tobacco	Indices of Industrial Production	1.88
Services	Food services	Monthly Survey on Service Industries	11.28
	Transportation	Indices of Tertiary Industry Activity	5.56
	Medical and other health care services	Indices of Tertiary Industry Activity	5.00
	Services for amusement and hobbies	Indices of Tertiary Industry Activity	4.79
	Travel services and accommodations	Monthly Survey on Service Industries, etc.	1.97
Total			72.93

Note: Figure for "transportation" is a weighted average of the figures for "railway," "bus," "taxi," and "air," calculated using the CAI weights.
Source: Bank of Japan.

4. Results

This section examines how well the ALC fits the CAI. The alternative data available for analysis span the period from January 2018 to April 2021. Using these data, we intend to nowcast the CAI for April 2021. To analyze the ALC's utility in greater depth, we also conduct "real-time forecasts" where we consider a hypothetical situation in which the data are accumulated on a monthly basis beginning in January 2019. Then, we perform nowcasting using the data available up until that month, although the results should be interpreted with caution due to the data's limited sample period.[10]

4.1. Item-level ALC

First, upon examining each movement, we notice that all three types of alternative data, which we use to calculate the ALC, appear to accurately capture developments in each item

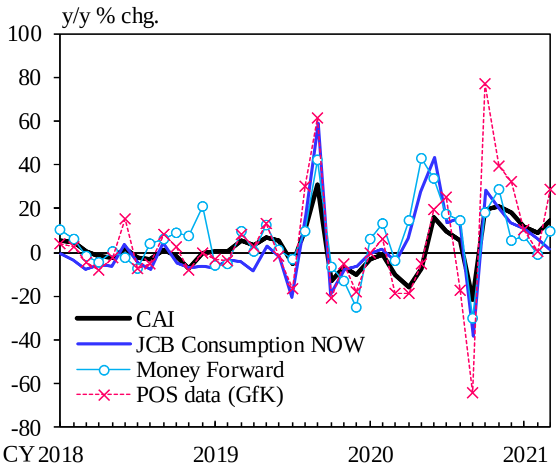
(5.98), communications (5.18), automobiles (4.12), financial services (1.83), care services (1.82), automobile maintenance (1.62), ceremonial occasions (1.38), supplementary tutorial schools (1.33), automobile parking (1.26), non-life insurance (1.16), public broadcasting (0.38), books and magazines (0.32), newspapers (0.36), game software (0.23), and postal services (0.10). The figures in parentheses indicate the weights of each item in the CAI.

[10] We cannot reconstruct the exact same situation because each dataset is revised retroactively. Thus, the examination in this section is to be called "pseudo-real-time forecasts."

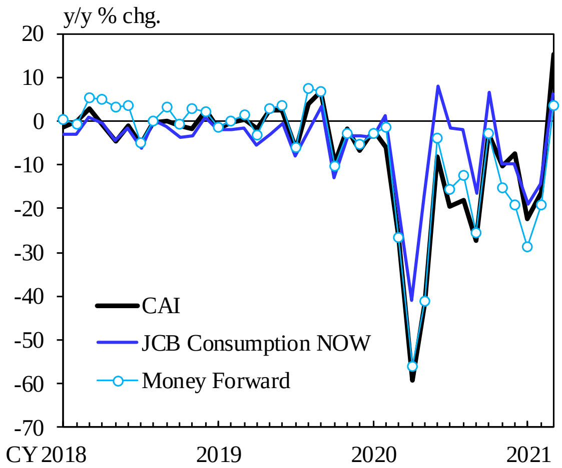
in the CAI; this suggests that these alternative data are useful in capturing private consumption broken down by item (Chart 6).

Chart 6. Comparison of the CAI and the Alternative Data (by Item)

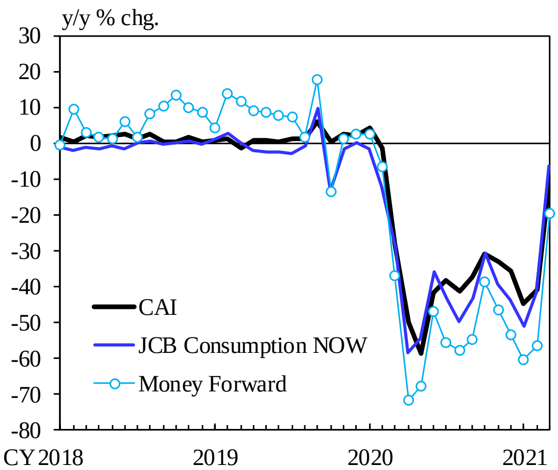
(a) Household electrical appliances



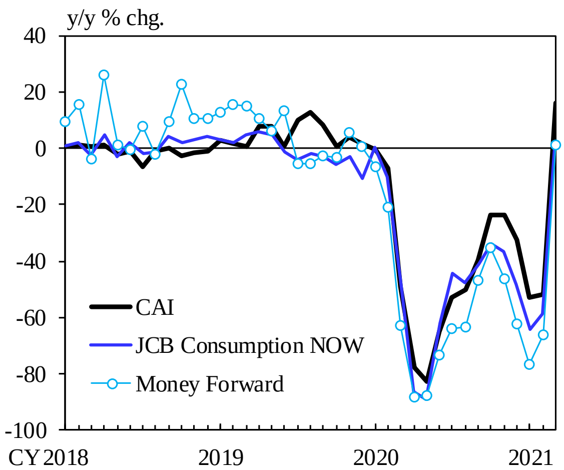
(b) Clothes



(c) Transportation



(d) Travel services and accommodations



Note: Figures for "transportation" are the weighted average of those for "railway," "bus," "taxi," and "air," calculated by using the weights of the CAI.

Sources: Nowcast Inc./JCB, Co., Ltd.; Money Forward, Inc.; Ministry of Economy, Trade and Industry; Bank of Japan; etc.

We examine the fit between the ALC and the CAI by item on this basis. For "household electrical appliances" in the CAI, POS data (GfK) most accurately capture its movements, such as the decline after the spread of COVID-19 and subsequent pick-up. "Clothes" in the CAI and the corresponding MFI are well matched, whereas "transportation" and "travel services and accommodations" in the CAI and the corresponding series in the *JCB Consumption NOW* are also well matched. In particular, GfK excels at capturing "household electrical appliances" in the CAI because it encompasses a diverse range of products regardless of payment method (although limited to the purchases at actual stores). Meanwhile, the MFI accurately captures developments in "clothes" in the CAI because it includes both online credit card purchases and in-store cash purchases. *JCB Consumption NOW* accurately captures "travel services and accommodations" in the CAI because payments for this component are frequently made with credit cards. Furthermore, the fact that *JCB Consumption NOW* does not include outbound tourism consumption appears to be the reason for their favorable fit. (Because the MFI covers outbound tourism consumption, it was affected to a greater extent than the CAI by the decline in consumption related to travel services and accommodations during the pandemic; this appears to result in a gap between them.)

Chart 7 depicts the relationship between the ALC and the CAI source statistics for each item. The item-level ALC, that is, the series that has the best fit with the corresponding series in the CAI among the three types of alternative data, fits well with the source statistics, with correlation coefficients exceeding 0.8 for the majority of the items (Chart 7). These findings suggest that alternative data can be useful in capturing developments in the CAI's major source statistics.

Chart 7. Correlations between the Source Statistics of the CAI and the Item-Level ALC

Consumption type	Item	Source statistics	Correlation coefficient with the ALC
Durable goods	Household electrical appliances	Current Survey of Commerce	0.886
Non-durable goods	Food and beverages	Current Survey of Commerce	0.356
	Drugs, cosmetics, etc.	Current Survey of Commerce	0.735
	Clothes	Current Survey of Commerce	0.957
	Electricity, gas, and water	Survey of Electric Power Statistics, etc.	0.968
	Fuel	Current Survey of Commerce	0.973
	Tobacco	Indices of Industrial Production	0.879
Services	Food services	Monthly Survey on Service Industries	0.962
	Transportation	Indices of Tertiary Industry Activity	0.981
	Medical and other health care services	Indices of Tertiary Industry Activity	0.846
	Services for amusement and hobbies	Indices of Tertiary Industry Activity	0.971
	Travel services and accommodations	Monthly Survey on Service Industries, etc.	0.970

Note: Figure for "transportation" is a weighted average of the figures for "railway," "bus," "taxi," and "air," calculated using the CAI weights.

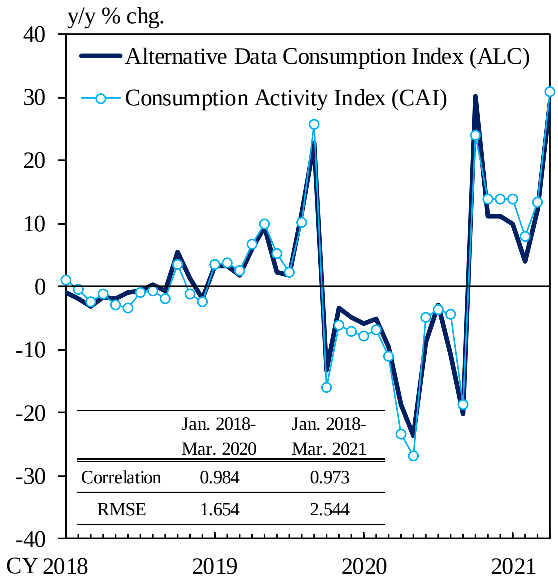
Sources: Nowcast Inc./JCB, Co., Ltd.; Money Forward, Inc.; GfK; Ministry of Economy, Trade and Industry; Bank of Japan; etc.

4.2. Macro level ALC

The ALC for durable goods, non-durable goods, and services, which are all calculated using the item-level ALC calculated in Section 4.1, accurately nowcasts the corresponding series in the CAI (Chart 8 (a)-(c)). In particular, the ALC for each consumption type accurately captures the large movements in consumption activity caused by the spread of COVID-19 since spring 2020. With respect to durable goods, the ALC closely reflects the front-loaded increase and subsequent decline in consumption prior to and after the October 2019 consumption tax hike. For all three consumption types, the correlation coefficient between the ALC and the CAI is greater than 0.9. We also use the root mean squared error (RMSE) to calculate the average errors of the ALC over the sample period. Although the RMSE of the ALC for durable goods is somewhat large (due to the large fluctuations in consumption of durable goods), the RMSE of the ALC for other consumption types is only around 1 percentage point. Moreover, the RMSEs of the ALC for all consumption types are higher for the sample period of January 2018 to March 2021, including the period when the significant impact of COVID-19 spread was observed, than for January 2018 to March 2020. However, the RMSEs for all consumption types are only around 2.5 percentage points at the maximum, indicating that the ALC performs well in capturing the CAI both before and after the COVID-19 outbreak.

Chart 8. Macro Level ALC

(a) Durable goods



(b) Non-durable goods

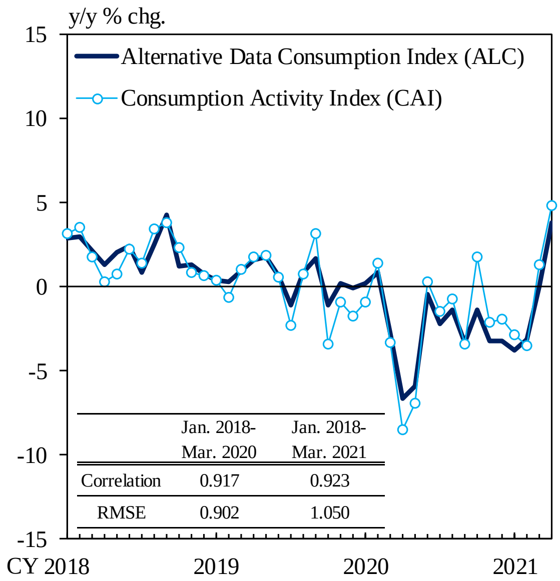
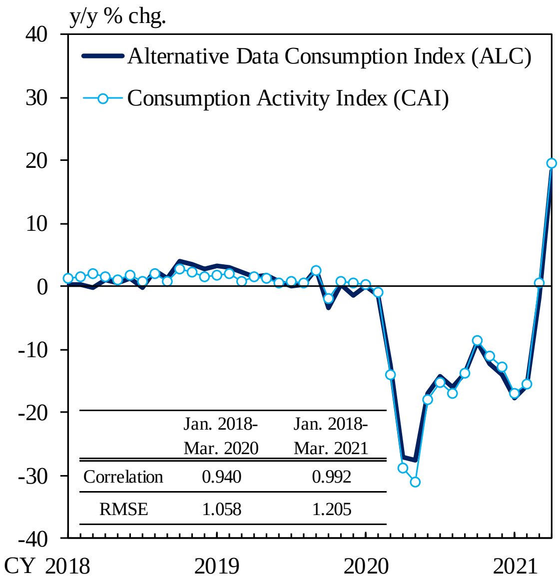
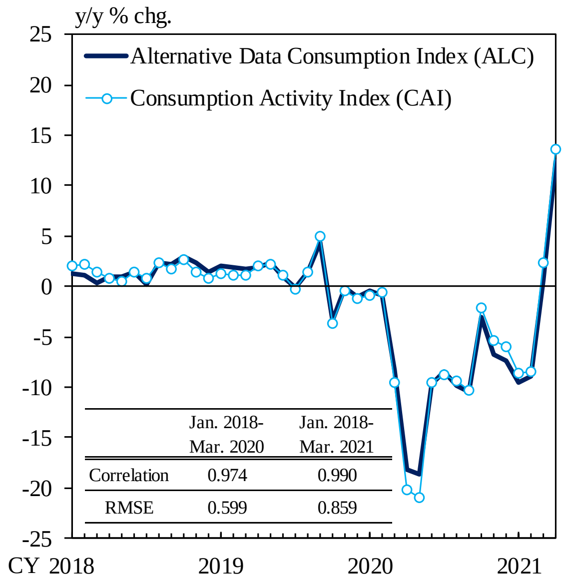


Chart 8. Macro Level ALC (Cont'd)

(c) Services



(d) All items



Sources: Nowcast Inc./JCB, Co., Ltd.; Money Forward, Inc.; GfK; Ministry of Economy, Trade and Industry; Bank of Japan; etc.

Regarding the ALC for all items, the correlation coefficient is much higher than 0.9, whereas the RMSE is below 1 percentage point. This indicates that the ALC for all items has an exceptionally good fit with the CAI over the sample period (Chart 8 (d)). In particular, the ALC accurately captures the considerable decline seen in April-May 2020 due to the impact of COVID-19 and the subsequent pick-up seen from summer to autumn 2021. These findings suggest that using alternative data allows us to accurately nowcast rapid changes in consumption (e.g., those seen during the pandemic) relatively early in the middle of the following month.

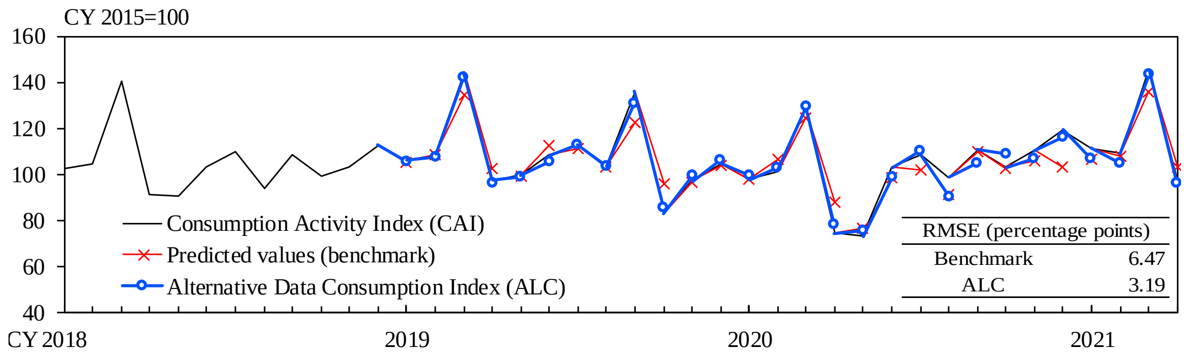
4.3. Assessment of the ALC based on real-time forecasts

Thus far, we have discussed the performance of the ALC based on an in-sample fit to past CAI data by estimating using the full sample (from January 2018 to March 2021). To examine the ALC's utility from a different perspective, we also conduct "real-time forecasts," by creating a hypothetical situation in which we nowcast the CAI every month as alternative data become available. We use the predicted values for each item covered in the CAI as benchmarks when evaluating the ALC's performance. The SARIMA model is used to calculate the predicted values. In this hypothetical scenario, we nowcast the CAI from January 2019 to April 2021. We assume that nowcasting is conducted in the middle of each month when the alternative data become available.

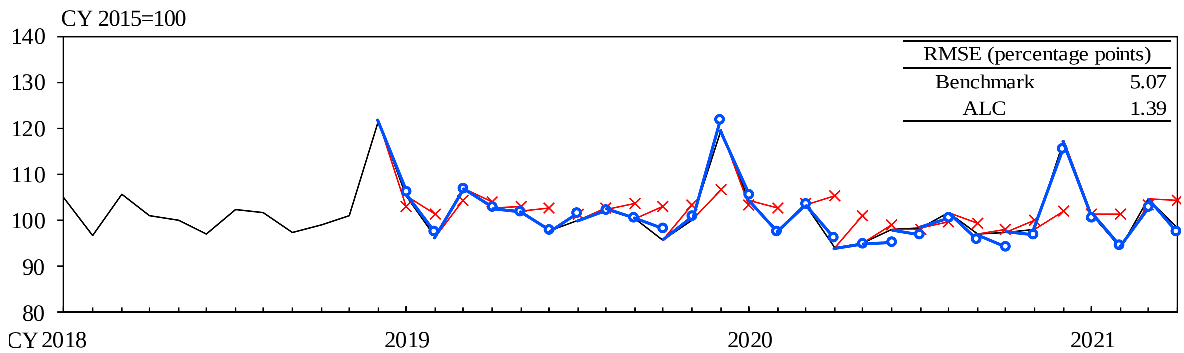
The real-time forecast results show that, for all three consumption types, the ALC nowcasts the CAI more accurately than the predicted values (benchmark) throughout the experiment (Chart 9). In particular, the ALC closely reflects the sudden drop in the CAI for services observed immediately following the outbreak of COVID-19 (around spring 2020), whereas the predicted values indicate a significant recovery, consistent with the past trend. The ALC for all items also accurately captures private consumption at the macro level, as evidenced by the ALC's forecast error (RMSE) being less than half that of the benchmarks.

Chart 9. Results of the Real-Time Forecasts

(a) Durable goods



(b) Non-durable goods



(c) Services

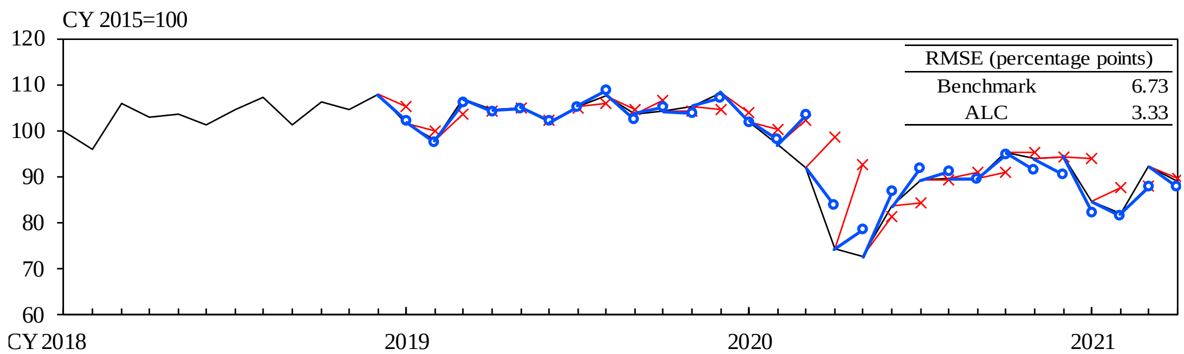
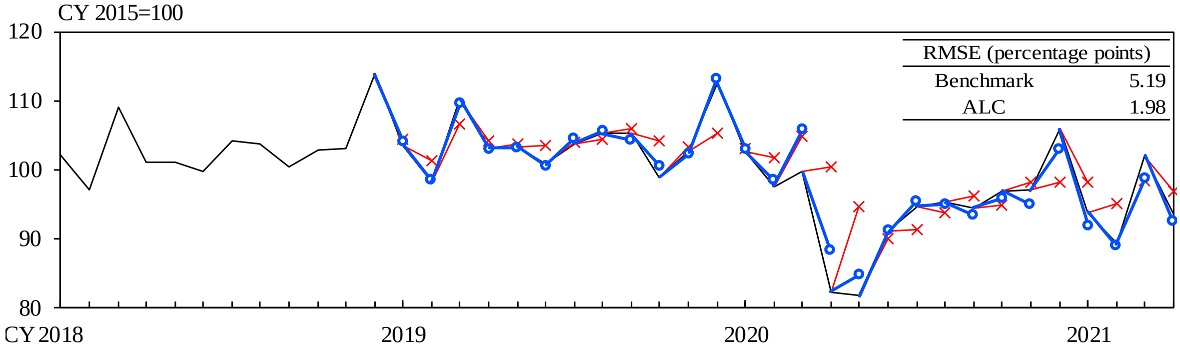


Chart 9. Results of the Real-Time Forecasts (Cont'd)

(d) All items



Note: The RMSEs are calculated based on the year-on-year rates of change. The sample period is from January 2019 to April 2021.

Sources: Nowcast, Inc./JCB, Co., Ltd.; Money Forward, Inc.; GfK; Ministry of Economy, Trade and Industry; Bank of Japan; etc.

Therefore, the ALC performs well in real-time forecasts overall. Based on these findings, we conclude that the ALC can be useful in practice, though the results should be interpreted with caution because the data sample period only extends from January 2018 to the present. (As of early 2019, the sample period for each item only lasted approximately 10 months.)

5. Conclusion

Using multiple sets of consumption related alternative data, we construct the ALC to nowcast developments in Japan's private consumption at the macro level. The ALC performs well in grasping consumption developments at the item or industry level and nowcasting consumption developments at the macro level. For example, it accurately captures the large fluctuations in consumption since spring 2020 due to the spread of COVID-19 approximately 3 weeks before the release of the CAI. Our analysis suggests that, to improve estimation accuracy when exploiting alternative data in nowcasting, the following two points warrant attention: (1) the definition or coverage of alternative data is not always the same as those of macro statistics, and (2) alternative data tend to include outliers and/or biases.

Although this study focuses on the timeliness of alternative data, future research could focus on its breadth and granularity. For example, *JCB Consumption NOW* and Money Forward data extensively cover online consumption of goods and services (e.g., e-commerce or consumption related to travel and accommodations taken place on online reservation websites), which have recently increased. Therefore, they could perform well in tracking a new trend in consumption when compared to existing statistics, whose data coverage is more limited to consumption expenditures at actual stores. Furthermore, detailed data by individual attribute can be used as pseudo-panel data, and thus, they may be functional in structural analysis of, for example, household behavior patterns. We believe that as alternative data accumulate and users' knowledge grows, more in-depth analyses will be possible. Working on such research and studies remains a future work.

References

Bounie, David, Youssouf Camara, and John W. Galbraith (2020), "Consumers' Mobility, Expenditure and Online-Offline Substitution Response to COVID-19: Evidence from French Transaction Data," *Econometric Modeling: Microeconometric Models of Household Behavior eJournal*.

Carvalho, Vasco M., Juan R. Garcia, Stephen Hansen, Álvaro Ortiz, Tomasa Rodrigo, José V. Rodríguez Mora, and Pep Ruiz (2021), "Tracking the COVID-19 Crisis with High-resolution Transaction Data," Royal Society Open Science, 8, pp. 210-218.

Chapman, James T. E., and Ajit Desai (2021), "Using Payments Data to Nowcast Macroeconomic Variables During the Onset of COVID-19," Bank of Canada Staff Working Paper, 2021-2.

Chikamatsu, Kyosuke, Naohisa Hirakata, Yosuke Kido, Kazuki Otaka (2021), "Mixed-Frequency Approaches to Nowcasting GDP: An Application to Japan," *Japan and the World Economy*, Vol. 57, 101056.

Dunn, Abe, Kyle Hood, Andrea Batch, and Alexander Driessen (2021), "Measuring Consumer Spending Using Card Transaction Data: Lessons from the COVID-19 Pandemic," *AEA Papers and Proceedings*, Vol. 111, pp. 321-325.

Dunn, Abe, Kyle Hood, and Alexander Driessen (2020), "Measuring the Effects of the COVID-19 Pandemic on Consumer Spending Using Card Transaction Data," BEA Working Paper Series, WP2020-5.

Kameda, Seisaku (2022), "Use of Alternative Data in the Bank of Japan's Research Activities," Bank of Japan Review Series, 2022-E-1.

Matsumura, Kohei, Yusuke Oh, Tomohiro Sugo, and Koji Takahashi (2021), "Nowcasting Economic Activity with Mobility Data," Bank of Japan Working Paper Series, No. 21-E-2.

Nakajima, Jouchi, Masato Takahashi, and Tomoyuki Yagi (2022), "An Assessment of Online Consumption Trends in Japan during the COVID-19 Pandemic," Bank of Japan Working Paper Series, No. 22-E-11.

Nakamura, Koji, Hiroshi Kawata, Masaki Tanaka, and Lisa Uemae (2016), "The Consumption Activity Index," BOJ Reports & Research Papers.

Nakazawa, Takashi (2022), "Constructing GDP Nowcasting Models Using Alternative

Data," Bank of Japan Working Paper Series, No. 22-E-9.

Takahashi, Masato, Takuji Kondo, Ko Munakata, Tomohiro Okubo, and Yuto Iwasaki (2021), "Revision of the Consumption Activity Index to Capture Recent Changes in Consumption Patterns," BOJ Reports & Research Papers.

Urasawa, Satoshi (2022), "GDP Nowcasting: Achievements and Challenges-Towards the Utilization of Alternative Data," Presentation prepared for the study group on economic models using AI and big data, February 4th, 2022.