IFC-Bank of Italy Workshop on "Data Science in Central Banking: Applications and tools"

14-17 February 2022

Nowcasting economic activity with mobility data[1]

Koji Takahashi, BIS,
Kohei Matsumura, Yusuke Oh and Tomohiro Sugo, Bank of Japan

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Nowcasting Economic Activity with Mobility Data*

Kohei Matsumura†; Yusuke Oh‡; Tomohiro Sugo§; Koji Takahashi**

Abstract

We develop high frequency indexes to measure sales in service industries and production activity in the manufacturing industry by using GPS mobility data from mobile applications. First, focusing on the possibility that the number of customers in service industries can be estimated using mobility data, we develop indicators to capture economic activity in amusement parks, shopping centres, and food services. We show that using GPS mobility data, it is possible to nowcast economic activity in the service industries, in real time, with a high level of precision---something which conventional statistics are largely unable to assist. In addition, by using a clustering method, we can construct an indicator with even better nowcasting performance. Second, in the manufacturing sector we identify the locations of large factories using factory-level data from the Economic Census and by utilizing hourly and daily mobility patterns such as a daytime ratio. We then construct indicators for nowcasting production based on the population in the specified areas. We find that we can nowcast production with a high level of precision for some labour-intensive industries including the transportation equipment and production machinery industries. These results suggest that mobility data are a useful tool for nowcasting macroeconomic activity.

Keywords: C49, E23, E27

JEL classification: mobility data, nowcasting, clustering

* The authors are grateful to Seisaku Kameda, Kenji Sakuta, Takuji Kawamoto, Kenichi Sakura, Kazushige Kamiyama, Jouchi Nakajima, Tomoyuki Iida and participants of the Big Data forum co-hosted by the University of Tokyo Center for Advanced Research in Finance and the Bank of Japan Research and Statistics Department, for helpful comments and discussions. Any remaining errors are ours. 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 Bank for International Settlements.

† Bank of Japan, kouhei.matsumura@boj.or.jp

‡ Bank of Japan, yuusuke.ou@boj.or.jp

§ Bank of Japan, tomohiro.sugou@boj.or.jp

** Bank for International Settlements, koji.takahashi@bis.org

1 Introduction

Since the 2010s, the rapid development of information and communication technology has made it possible to collect and use big data in business, which were previously unobtainable from conventional statistics and questionnaire surveys. In particular, in business marketing, big data based on the global positioning system (GPS) of cellular phones are widely used as a source to understand the consumption behavior of customers by taking into account their characteristics. Meanwhile, in the field of economics, especially in macroeconomics, point of sale (POS) data have been widely used in the analysis of supermarket sales and prices since the 1990s. The use of big data in macroeconomics has been rapidly spreading since the beginning of 2020, when the COVID-19 pandemic began to gather pace.

More specifically, the spread of COVID-19 has generated large economic fluctuations in short periods due to subsequent lockdowns and declarations of a state of emergency. The economic turmoil brought on by the pandemic has, therefore, massively increased the importance of understanding economic conditions in a timely manner for policy makers. However, the conventional statistics and questionnaire surveys on which central bankers have relied for economic analysis take at least several weeks for data to be released after the survey. This is due to the time which the process of collecting and compiling the data takes. To address this issue, the use of big data, especially high frequency data, has been rapidly spreading among central bankers as well as the utilization of information obtained through interviews with firm managers.[1]

In this paper, focusing on GPS data, we show how to use big data to nowcast economic conditions in real time from macroeconomic perspectives. Specifically, by combining the GPS data with information on coordinates of commercial and public facilities such as shops and factories, we closely examine which sectors the data can be applied to for nowcasting with a high level of precision and frequency. We find that for some sectors, we can nowcast household consumption and firm production with high accuracy by extracting related information from the mobility data. This result implies that mobility data are useful not only for marketing but also for understanding macroeconomic conditions.

The remainder of this paper is organized as follows. Section 2 explains related literature and the dataset used in this paper. Section 3 shows that the GPS data are useful for nowcasting service industries. Section 4 demonstrates a methodology to nowcast the index of the industry production. Section 5 concludes by pointing out the caveats of using mobility data for these purposes.

2 Literature Review and Data

In this section, we discuss literature related to our paper and explain our dataset.

[1] In this paper, big data include (1) high frequency data that are updated frequently, (2) high granular data such as transaction data between companies, and (3) text data such as those on social networking services.

2.1 Literature Review

Our paper is mainly related to two strands of literature. First, this paper relates to the literature on analyses of human mobility based on smartphone GPS data. Since the beginning of the COVID-19 pandemic, a growing number of papers study the effect of lockdowns using mobility data. For example, in the United States, Couture et al. (2021) develop a location exposure index and a device exposure index using smartphone mobility data. The former describes county-to-county movements of people and the latter quantifies people's exposure to others in commercial venues. Furthermore, using smartphone GPS data, Coven and Gupta (2020) show that since the pandemic, people with higher incomes tend to move away from urban areas whereas people with lower incomes are likely to continue going out and commuting to their work places as before.[2] On China, Fang et al. (2020) investigate the effects of lockdown on human mobility using smartphone GPS data. On Japan, Watanabe and Yabu (2020) develop a "stay-at-home" index and quantitatively investigate to what extent Japanese shelter-in-place behavior is explained by intervention effect and information effect. The former represents a direct effect resulting from an intervention such as the declaration of a state of emergency whereas the latter means an effect resulting from an announcement such as a news story about the number of infected people. While these studies focus on the mobility of people, they do not explicitly investigate the relationship between mobility data and economic activity.

The second strand of related literature is studies on the development of nowcast indexes. Among them, Cajner et al. (2019), together with other economists of the Federal Reserve Board (FRB), develop the ADP-FRB active employment index to understand the labor market conditions using data from a private company, Automatic Data Processing (ADP). The FRB uses the index in order to grasp current labor market conditions and wage. In addition, a growing number of studies use payment data for nowcasting. For example, Aprigliano et al. (2019) show that Business-to-Business and Business-to-Consumer payment data help to improve the accuracy of nowcasting GDP, business fixed investment, and consumption. Furthermore, Galbraith and Tkacz (2018) find that payment data for debit cards and checks are useful for nowcasting GDP and consumption and Aladangady et al. (2019) develop an index to nowcast consumption in real time by exploiting debit and credit card payment data. Since the occurrence of the pandemic, Chetty et al. (2020) developed economic indexes using big data compiled by private companies. They report that based on the indexes, a reduction in the service consumption of people with higher incomes leads to a decrease in income and the consumption of people with lower incomes working in those industries. These approaches are useful for practitioners as they allow them to grasp macroeconomic conditions before related public statistics are released. However, these studies use high frequency data that are directly linked to economic activities and do not use location data.

As for the studies on nowcasting economic activity using location data, Dong et al. (2017) develop indexes to nowcast sales of firms and consumption in some service industries in China. In addition, Arslanalp et al. (2019) and Cerdeiro et al. (2020) develop indexes to nowcast trade volume in the world using the traffic data of vessels.

[2] Using smartphone mobility data, Chen and Pope (2020) find that people with higher incomes travel longer distances and go to many places in the United States.

However, to date, few studies use location or mobility data for nowcasting economic activity.

Our paper extends these two bodies of literature and thereby contributes to the development of nowcasting indexes of economic conditions with high frequency and a high level of precision, which indicates that mobility data are a useful tool for nowcasting macroeconomic activity.

2.2 Data

We use mobility data from January 2017 to March 2020 compiled by Agoop. The data show the estimated number of people in every hour in each mesh element, which is defined as a 100m$\times$100m square, dividing Japanese territory into about 20 million squares.[3] The estimation is based on GPS data, which are collected through smartphone applications with the approval of users.

However, we cannot measure any economic activities such as consumption and production simply using hourly population data because the data only include information on the coordinates of meshes in addition to hourly population. Therefore, in order to associate the number of people in a mesh with a specific economic activity, we combine the data with the 1) Economic Census for Business Activity compiled by the Ministry of Economy, Trade and Industry, 2) National Land Numerical Information compiled by the Ministry of Land, Infrastructure, Transport and Tourism and/or 3) point of interest data with the application programming interface (API). Thus, we develop indexes to capture the consumption or production activity based on the number of people in a mesh.[4]

More specifically, we identify a type of a facility and building located in each mesh and infer whether the hourly population in the mesh is associated with consumption such as retail, leisure, or food services, or with production in factories. By doing so, we capture the economic activity based on the number of people in each mesh.

2.3 Calculation of Indicators

We develop an economic indicator from GPS data (EIG) to capture economic activity such as sales in service sectors using population data by following the steps below.

First, we specify nowcasting sector J and select a set of meshes (I_t^J) related to the sector in time t. Then, we sum the number of people (Pop_{it}) in each selected mesh i using weight w_{it}, which is calculated with other data such as the Economic Census for Business Activity. Thus, we obtain the total population in the selected meshes ($TotalPop_t^J$) as follows,

$$TotalPop_t^J = \sum_{i \in I_t^J} w_{it} \times Pop_{it}. \tag{1}$$

[3]The data do not include meshes such as mountainous areas where the average population is almost zero.

[4]The Economic Census for Business Activity and API point of interest data include information on the business categories of firms in their specific meshes. The National Land Numerical Information provides information on the land use, which allows us to relate the number of people in some meshes to a specific economic activity.

Second, as set I_t^J does not necessarily include meshes that cover all the facilities and buildings related to the industry, we define EIG as a normalized index as follows,

$$EIG_t^J = \frac{TotalPop_t^J}{TotalPop_s^J} \times 100 \tag{2}$$

where $TotalPop_s^J$ indicates the total number of people at reference point of time s. It should be noted that the number of people in meshes are available on an hourly basis. Therefore, economic activity can be analyzed hourly or weekly, as well as on the monthly frequency used in typical conventional statistics.

3 Mobility Data and Service Industries

This section illustrates the methodology of constructing indexes to nowcast current economic conditions in service industries. In service industries, the number of people in a commercial facility is likely to represent the number of its customers. Therefore, if we count the number of people accurately, we can nowcast the number of customers or sales in that facility.

However, it is unusual that only one type of economic activity is conducted in a mesh. Suppose we are interested in food service industry. However, a mesh where a restaurant is located often includes a supermarket or a department store. It is therefore important to exclude some meshes that are affected by customers of the department store and the supermarket in order to nowcast economic activity in the food service industry accurately.

Against this background, in what follows we develop indexes for nowcasting economic activity in amusement parks, shopping centers, and food service industry by using a statistical method to get rid of noise if necessary and we investigate the plausibility of the indexes by comparing them with existing conventional statistics.[5]

3.1 Amusement Parks

The population in amusement parks is expected to be a proxy for the number of visitors. In addition, there are not many large amusement parks in Japan, which allows us to choose meshes that cover almost all amusement parks. We sum the number of people each day in meshes that cover the main amusement parks.

Figure 1 shows the hourly population by day of the week, indicating that the number of visitors is relatively high in the daytime on weekends as well as on Mondays and Fridays, which correspond to the days before and after weekends. This figure shows that the amusement park characteristic that visitors increase on weekends is captured well by the indicator. Figure 2 shows the year-over-year rate of change of the population in specific meshes and it appears to track the change in the number of visitors to amusement parks obtained from Current Survey of Selected Service Industries, which indicates that the GPS data are useful for nowcasting official statistics.

[5]As Chetty et al. (2020) pointed out, when we use alternative data for economic analysis, we need to modify biases arising from the system changes of data providers. Following the literature, we also modified discontinuity that arises from changes in the data collection process of the data provider.

3.2 Shopping Center

As a shopping center (SC) usually has a large site area, it is easy to extract meshes most related to consumption in SCs by excluding meshes where the hourly population is affected by other economic activities. First, we extract meshes that include the addresses of shopping centers listed by the Japan Council of Shopping Centers (3,203 meshes). Then, we exclude meshes including stations because the hourly population in stations is affected by many other economic activities rather than simply consumption in shopping centers. We also exclude meshes where the average population on weekends is smaller than that on weekdays, in order to exclude meshes with facilities such as offices. After this process, the number of selected meshes will be 2,361 meshes.

Figure 3 shows the average hourly population by day of the week in the selected meshes, indicating that the number of visitors increases in the afternoon on weekends and at 6 p.m. on Fridays. This pattern implies that the population captures the number of visitors to the shopping centers well. In fact, as shown in Figure 4, the year-over-year growth rate of the population from 10 a.m. to 8 p.m. shows similar patterns with those of sales at shopping centers, which are collected by Japan Council of Shopping Centers. The result shows that mobility data are useful for capturing consumption. However, we should note that the index based on the population does not capture fluctuations before and after the consumption tax rate hike in October 2019. This suggests that mobility data may not capture changes in the consumption value per person.

3.3 Food Service Industry

In the food service industry, a site area of each restaurant is expected to be relatively small in general, compared to the size of a mesh. In addition, commercial facilities such as pubs are located in areas where various different activities are conducted. This complicates the selection of meshes in the food service industry. Given these characteristics, we take the following three steps to identify meshes whose population represents activity in the food service industry. First, following Mizuno et al. (2020), in order to exclude residential areas, we extract meshes where the daytime population is greater than 80% of nighttime population (27,595 meshes). Second, we count the number of restaurants within a 300m radius from the center of each mesh and choose meshes where the number is greater than 300.[6] In addition, we exclude meshes that include stations. The number of extracted meshes then becomes 758. Third, we categorize those meshes into five clusters by using a k-medoids method and then exclude clusters of meshes whose hourly population appears to be affected by other economic activities.[7]

Figure 5 illustrates the hourly population by cluster on weekdays, indicating that cluster 1 (435 meshes) has peaks at noon and 7 p.m. This pattern suggests that the population in cluster 1 successfully captures the visitors to restaurants. On the other hand, the population in other clusters has a peak at 3 p.m., which implies that the

[6]We use Gurunavi API to count the number of restaurants within a 300m radius from the center of each mesh.

[7]k-medoids is a partitioning method of clustering that splits the data points into k clusters by choosing a medoid of a cluster so that its average dissimilarity to all the objects in the cluster is minimal.

population in those meshes mainly includes workers in offices. As shown in Figure 6, the year-over-year changes in the population of cluster 1 and the consumption index in food services from JCB Consumption NOW show similar fluctuations over time. In addition, the correlation between them is higher than the correlation with the population before clustering as shown in Table 1. Note that even if we exclude the data in March 2020 when the consumption sharply declined because of the shortened hours of restaurants due to the pandemic, the correlation remains high. This indicates that the indicator successfully captures consumption in the food service industry even in normal times.

For the above analysis on the three sectors in the service industry, some may point out that after the pandemic, the number of visitors to commercial facilities in service industry has drastically declined, which increased the correlation between sales and the population. Therefore, some may claim that a high correlation in turbulent times does not guarantee high nowcasting performance in normal times. However, we only use data up to March 2020 when the effect of the pandemic on consumption was limited and show that even before the serious deterioration of the pandemic, the indicator demonstrates a high correlation with other statistics. This result implies that the indicators are a useful tool for nowcasting economic conditions in a timely manner not only in turbulent times when the economic situation could suddenly change, but also in normal times.

4 Mobility Data and Manufacturing Industry

In this section, we outline our methodology for using mobility data for nowcasting industrial production. Generally speaking, production in the manufacturing industry is determined by many different factors—including the utilization rate of capital and level of technology progress—and the impact of each factor differs across sectors. However, irrespective of other factors, labor input is thought of as an important factor for almost all sectors by economists.

If production can be approximated by a simple function of labor input and if the labor input can be captured by population in factories, an indicator based on the population would be a useful tool for nowcasting production and judging economic conditions.

As with the service industry discussed in the previous section, provided that the hourly population in a factory can be a proxy for labor input there, we develop indicators for industrial production. In this section, first, we show a relationship between the industrial production and labor input of conventional statistics. Then, we explain the methodology to construct indicators for nowcasting the industrial production.

4.1 Relationship between Labor Input and Production

Following the discussion above, we check the relationship between labor input and production before investigating the relationship between the population and production. Table 2 shows the correlations between year-over-year growth rates of labor input (hour $\times$ number of workers) from the Monthly Labour Survey and the index of the industrial production, indicating that the correlations differ across sectors and they are high especially in the transportation equipment and the production machinery industries. For those industries, due to the strong linearity of the relationship between production and labor input, the population data are expected to be useful for nowcasting the production if we can successfully capture the labor input by the population. On the other hand, for other sectors including the electronic parts, devices and electronic circuits industry and the food, beverages and tobacco industry the correlations are almost zero due to the progress of automation in factories in these industries. For those sectors, the labor input is not a major driving factor in determining the production level and therefore it would be difficult to nowcast production in those sectors using mobility data.

4.2 Identifying Meshes Representing Location of Factories

By using panel data from the Economic Census for Business Activity, we specify meshes where production takes place. More specifically, we associate factory information, including addresses and value-added production from the Economic Census for Business Activity with meshes and sum the population in those meshes. In addition, to reduce noise, we use only the top ten thousand factories in the manufacturing industry in terms of value-added production among factories with a site area larger than 10,000 m^2.

The main issue for identifying meshes related to the production activity is that most of the factories have a site area larger than the minimum size of a mesh (100m by 100m square), but the census does not provide information on the shape of the factory. Nevertheless, in addition to the meshes corresponding to the address of factories in the panel data, we also need to choose additional meshes that sufficiently cover factories in order to nowcast production with a high level of precision. To address this issue, we employ a heuristic approach. After collecting meshes which sufficiently cover all areas that are expected to include factory sites, we exclude the less relevant meshes to production activity. Specifically, we choose meshes based on the three criterion: Sunday ratio, daytime ratio and average population. First, we use the Sunday ratio and daytime ratio as we think those ratios are useful measures to distinguish factories from residential areas. In fact, as shown in the upper panel of Figure 7, the Sunday ratio is low and the daytime ratio is high in a typical factory area. On the other hand, as shown in the lower panel of Figure 7, the Sunday ratio is high and the daytime ratio is low in a typical residential area. Thus, the Sunday ratio and daytime ratio should be good measures to separate factory areas from residential ones. In addition, by excluding meshes where the population is lower than specific thresholds, we mitigate the effects of noise arising from a small sample problem.

In this paper, considering the possibility that the meaningful thresholds of those ratios differ across sectors, we choose a combination of thresholds for the ratios so that the correlation between labor input and the population will be the highest. For example, in the transportation equipment industry, the correlation achieves the highest value of 0.86 when we set the Sunday ratio as 0.5 and the daytime ratio as 0.6 and the average population as 10 people, as shown in Table 3. Therefore, we use those values as the criterion for the transportation equipment industry. Even though the correlation between the population and the index of the industrial production differs across sectors, the population in the selected meshes can capture the labor input on average, as shown in Figure 8.

4.3 Activity Indicator Based on the Hourly Population Data

In this subsection, we explain the methodology to construct a proxy for production activity using population in the selected meshes. When we aggregate the population in these meshes, we should note that it does not necessarily provide the best nowcast to sum population over all hours and days. This is because the hourly population can change even as a result of an increase in the number of workers irrelevant to production, such as facility maintenance workers. Taking this point into account, we investigate all combinations of choices of (1) whether we include holidays and weekends and (2) which hour we aggregate the population data with. Then, we figure out a combination that provides us with the highest correlation between the indicator based on the population and the industrial production index. Table 4 shows that the population in the evening is an appropriate measure to capture the industrial production on average although there is heterogeneity across sectors.

Figure 9 shows the sum of the population for the selected window of hours in the identified meshes, indicating it traces the development of year-on-year changes in the index of the industrial production well. In particular, in October 2019 when an enormous typhoon hit Japan, the indicator based on the population has massively declined in tandem with the industrial production index, which suggests that we can nowcast the production with a high level of precision using mobility data.

Figure 10 shows the indicators based on the population by industry, indicating that the indicators for the transportation equipment and the production machinery industries nowcast the corresponding indexes of the industrial production relatively well. The high nowcasting performance is due to the fact that labor can explain production because those industries both have a high level of labor intensiveness and the population can serve as a good proxy for labor input. On the other hand, the indicator based on the hourly population for the electronic parts and devices industry shows a poor nowcasting performance, especially in 2019 when production substantially declined. This is because the industry is a capital intensive sector and the population only poorly tracks the labor input.

4.4 Extension: High-frequency Analysis

One possible extension of the analysis in the previous section is to enhance the frequency of the indicator. Since the original data is hourly, it is possible to evaluate the current economic situation in a more timely manner by changing the frequency of the indicator to weekly or daily.

For example, Figure 11 shows the weekly production indicator, which is seasonally adjusted by a statistical method, based on the population data for the transportation equipment industry.[8] The figure demonstrates that the indicator sharply dropped in the week of October 20th, 2019 in tandem with the industrial production index, which suggests that the indicator based on the population can capture the effect which the number 19 typhoon in 2019 had on the production activity in a timely manner. This result implies that mobility data are also helpful for capturing a sudden fluctuation of

[8] We use an open source software, called "Prophet," developed by Facebook for the seasonal adjustment.

economy even when the economy is hit by natural disasters such as heavy rains or typhoons.

5 Conclusion

In this paper, we develop indicators to capture sales in the service industries and production activity in manufacturing industries using the hourly population based on smartphone mobility data.

As the hourly population in a commercial facility correlates with the number of customers in the service industries, we develop activity indicators for amusement parks, shopping centers, and the food service industry. We find that the indicators based on population are useful for nowcasting activity in the service industry for which conventional statistics are unable to give an understanding of economic activity in real time. In addition, even for a sector where it is difficult to get rid of noise, we can construct an indicator with high nowcasting performance by using a statistical method such as clustering. Furthermore, in the manufacturing industry, we identify meshes related to production activity by using factory panel data from the Economic Census for Business Activity and by utilizing hourly and daily mobility patterns such as the daytime ratio. Then, we sum the population in specified meshes, thereby constructing indicators by industry for nowcasting. We find that we can nowcast with a high level of precision by using the indicators for labor intensive industries including the transportation equipment and production machinery industries.

These results suggest that mobility data are a useful tool for nowcasting macroeconomic activity.

We should note that the mobility data used in this paper include information on the hourly population in a 100m by 100m mesh but do not include information on characteristics such as age and gender of workers and consumers in the area. Therefore, we cannot analyze consumption behavior by attribute, such as that of elderly people, using mobility data alone. In addition, as the data used in this paper are collected through smartphones, they cannot reflect the behavior of people who do not use smartphones. Furthermore, the length of our data in terms of time series would be short to check the plausibility of our analysis. Moreover, as we combine the population data with the Economic Census in this paper, it is important to find other appropriate data for a targeted sector in order to construct an effective indicator. For example, if the relocation of a large factory from one of the selected meshes to another non-selected mesh occurred, the indicator would drastically drop. However, the reason for such a change cannot be found using population data alone. In such cases, we have to use other data in order to uncover the reason for the factory relocation and to control the effect on the indicator if needed. In addition, if we focus on production in a sector where the production process is highly automated, the utilization rate of capital and changes in productivity would be important determinants of production as well as labor input estimated by the population data. Disregarding these factors would degrade the accuracy of developed indicators. Investigating a methodology to address this issue will be a topic for our future research.

References

Aladangady, A., S. Aron-Dine, W. Dunn, L. Feiveson, P. Lengermann, and C. Sahm (2019): "From transactions data to economic statistics: constructing real-time, high-frequency, geographic measures of consumer spending," NBER Working Paper 26253.

Aprigliano, V., G. Ardizzi, and L. Monteforte (2019): "Using payment system data to forecast economic activity," International Journal of Central Banking, 15, 55–80.

Arslanalp, S., M. Marini, and P. Tumbarello (2019): "Big data on vessel traffic: Nowcasting trade flows in real time," International Monetary Fund Working Paper WP/19/275.

Cajner, T., L. D. Crane, R. A. Decker, A. Hamins-Puertolas, and C. Kurz (2019): "Improving the accuracy of economic measurement with multiple data sources: The case of payroll employment data," Finance and Economics Discussion Series 2019-065, Board of Governors of the Federal Reserve System.

Cerdeiro, D. A., A. Komaromi, Y. Liu, and M. Saeed (2020): "World seaborne trade in real time: A proof of concept for building AIS-based nowcasts from scratch," International Monetary Fund Working Paper WP/20/57.

Chen, M. K. and D. G. Pope (2020): "Geographic mobility in America: Evidence from cell phone data," NBER Working Paper 27072.

Chetty, R., J. N. Friedman, N. Hendren, M. Stepner, and the Opportunity Insights Team (2020): "How did Covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker based on private sector data," NBER Working Paper 27431.

Couture, V., J. I. Dingel, A. Green, J. Handbury, and K. Williams (2021): "Measuring movement and social contact with smartphone data: a real-time application to COVID-19," Journal of Urban Economics, 103328.

Coven, J. and A. Gupta (2020): "Disparities in mobility responses to Covid-19," NYU Stern Working Paper.

Dong, L., S. Chen, Y. Cheng, Z. Wu, C. Li, and H. Wu (2017): "Measuring economic activity in China with mobile big data," EPJ Data Science, 6, 1–17.

Fang, H., L. Wang, and Y. Yang (2020): "Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in china," Journal of Public Economics, 191, 104272.

Galbraith, J. W. and G. Tkacz (2018): "Nowcasting with payments system data," International Journal of Forecasting, 34, 366–376.

Mizuno, T., T. Ohnishi, and T. Watanabe (2020): "Visualization of the rate of self-quarantine based on big population mobiliy data," Artificial Intelligence (in Japanese).

Watanabe, T. and T. Yabu (2020): "Japan's voluntary lockdown," CARF Working Paper.

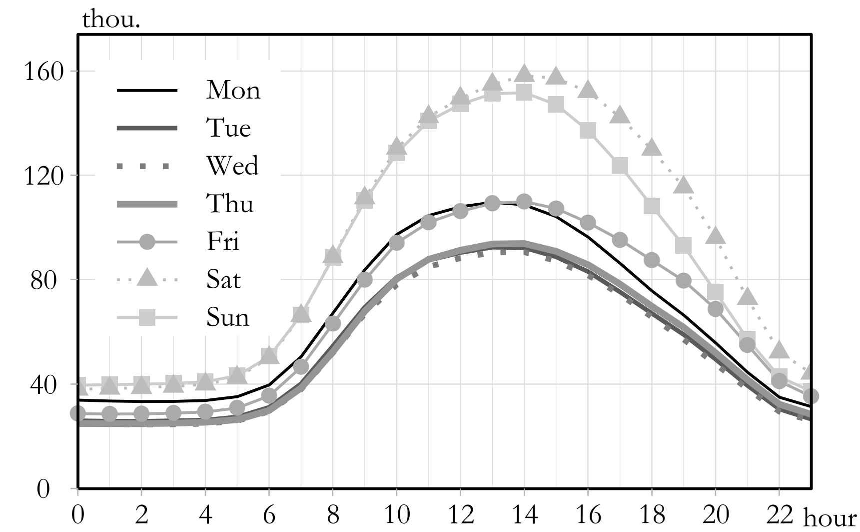


Figure 1: Hourly population by day of the week in amusement parks

Source: Agoop.

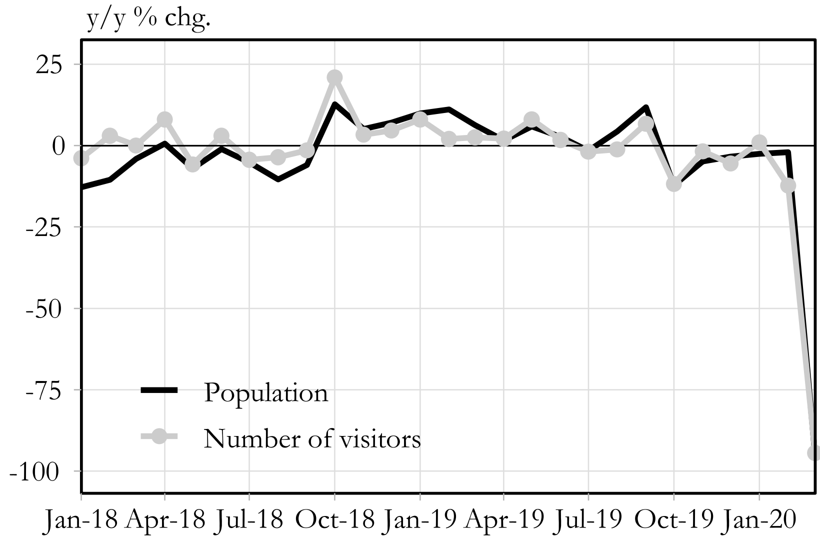


Figure 2: Indicator based on the hourly population in amusement parks

Note: "Number of visitors" indicates the number of visitors to amusement parks based on the Survey of Selected Service Industries.
Sources: Agoop; Ministry of Economy, Trade and Industry

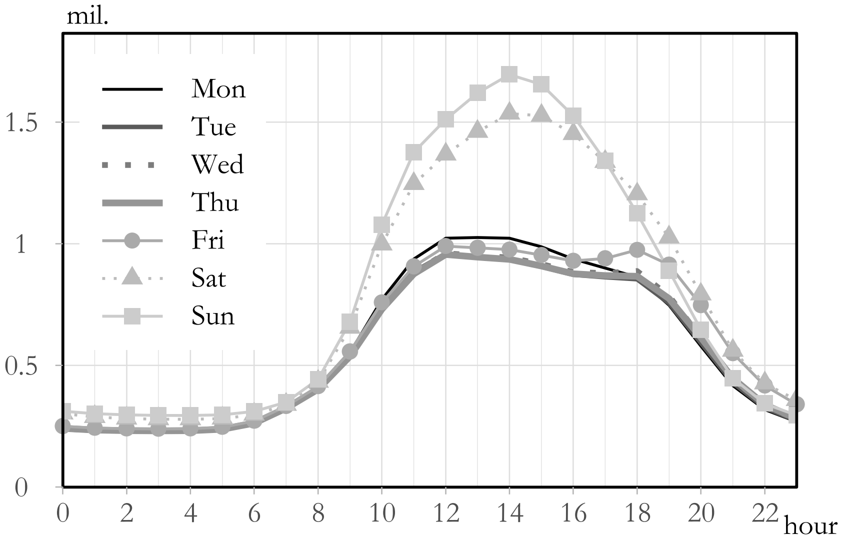


Figure 3: Hourly population by day of the week in shopping centers

Sources: Agoop; Japan Council of Shopping Centers; Ministry of Land, Infrastructure, Transport and Tourism.

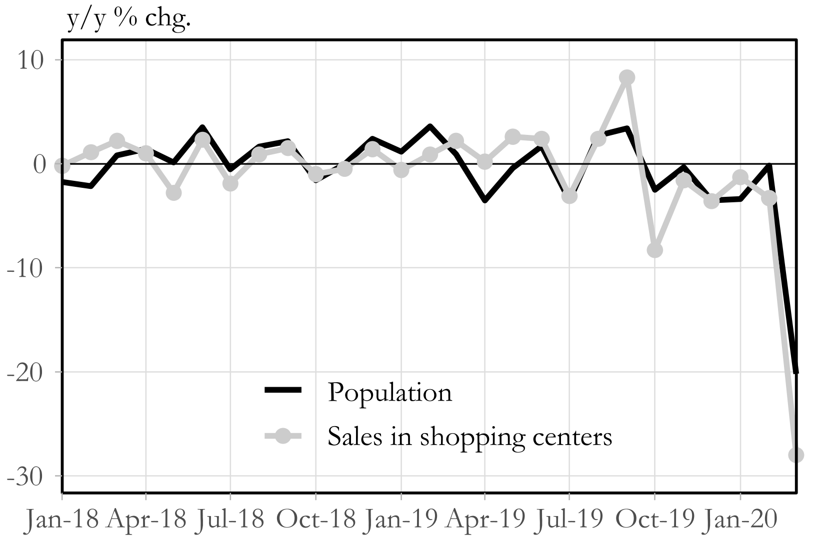


Figure 4: Indicator based on the hourly population in shopping centers

Note: "Sales in shopping centers" is based on the survey by Japan Council of Shopping Centers.
Sources: Agoop; Japan Council of Shopping Centers; Ministry of Land, Infrastructure, Transport and Tourism.

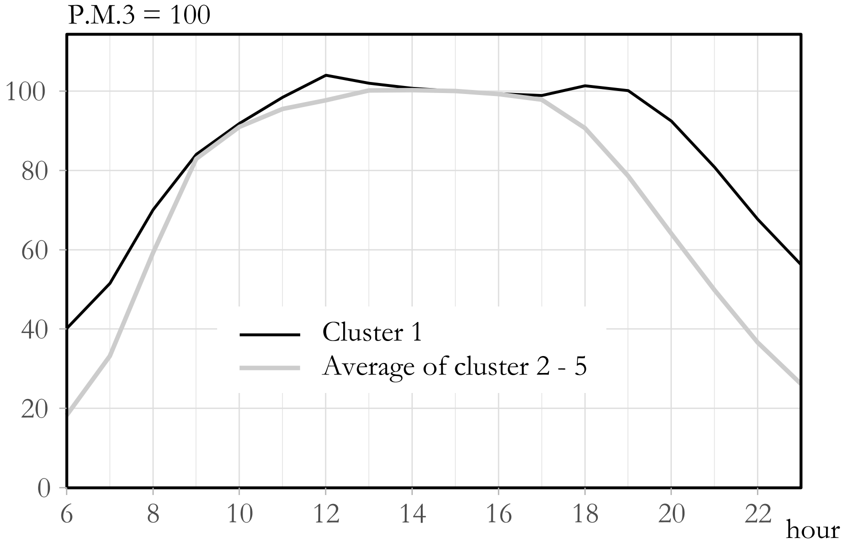


Figure 5: Hourly population on weekdays by cluster for food services

Sources: Agoop; Gurunavi; Ministry of Land, Infrastructure, Transport and Tourism.

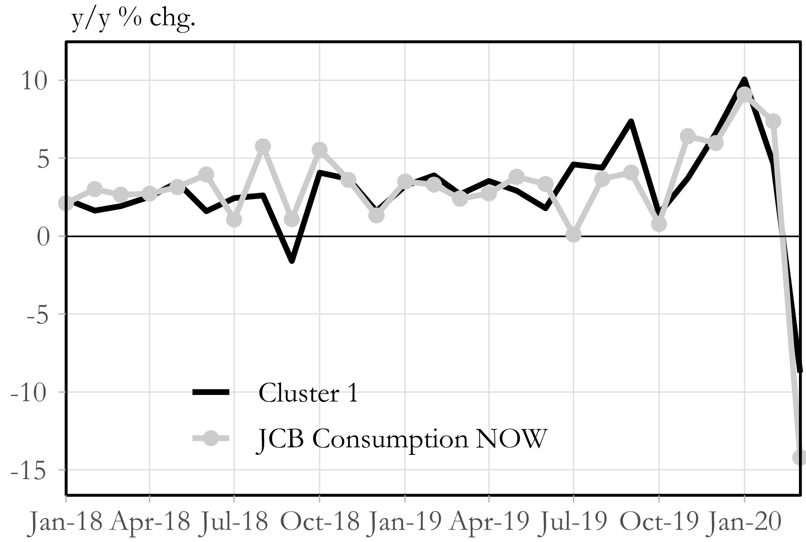


Figure 6: Indicator based on the hourly population for the food service industry

Note: "JCB Consumption NOW" indicates a consumption index for food services based on credit card transaction data.
Sources: Agoop; Gurunavi; Ministry of Land, Infrastructure, Transport and Tourism; NOWCAST, Inc./ JCB, Co., Ltd., "JCB Consumption NOW."

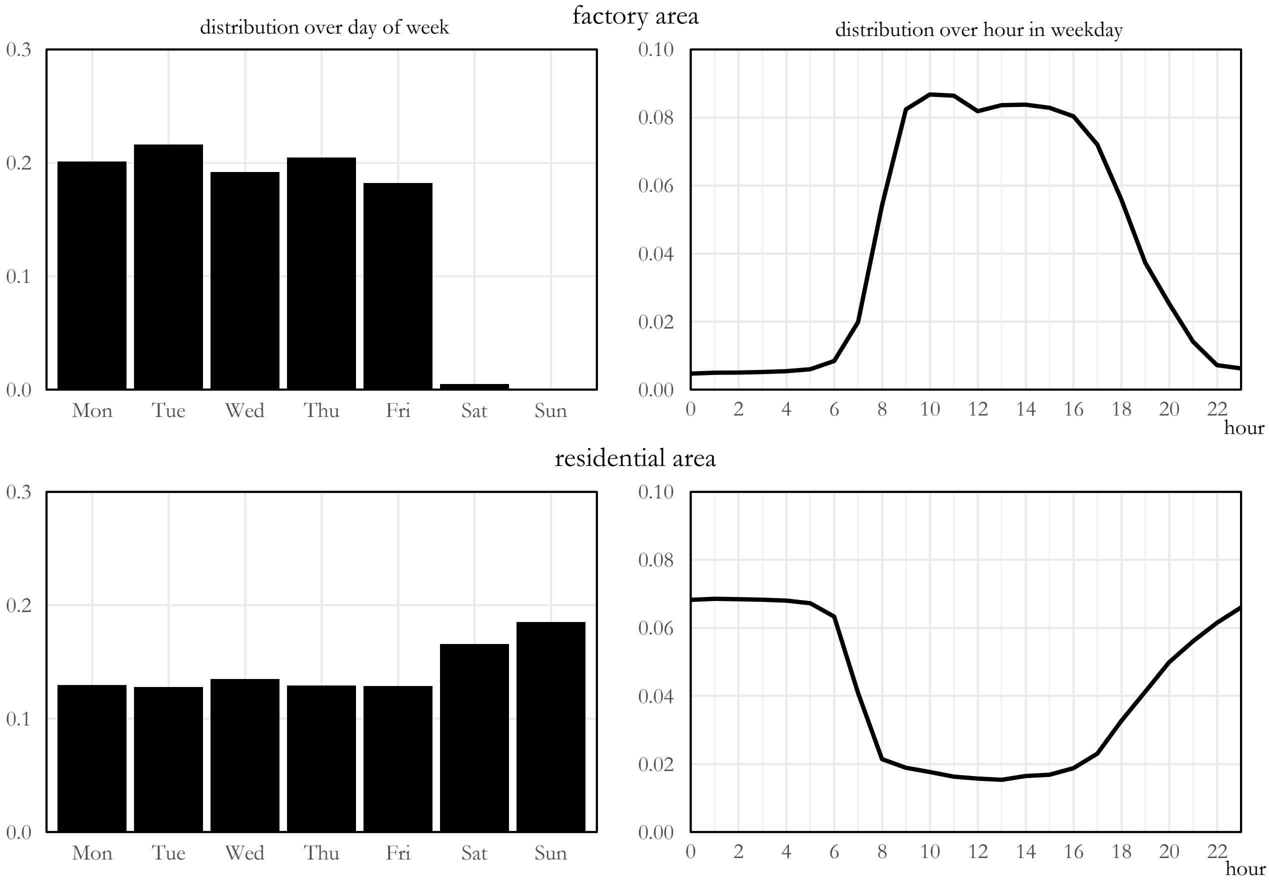


Figure 7: Hourly population in a typical factory area and residential area

$Note$: Figures show patterns of the hourly population in a typical factory area and residential area based on data on the weeks of which the previous and following weeks have five business days as well as those weeks themselves in order to eliminate the effect of public holidays. The left side panels show population by day of the week and the right side panels show average population by hour on weekdays.
$Source$: Agoop.

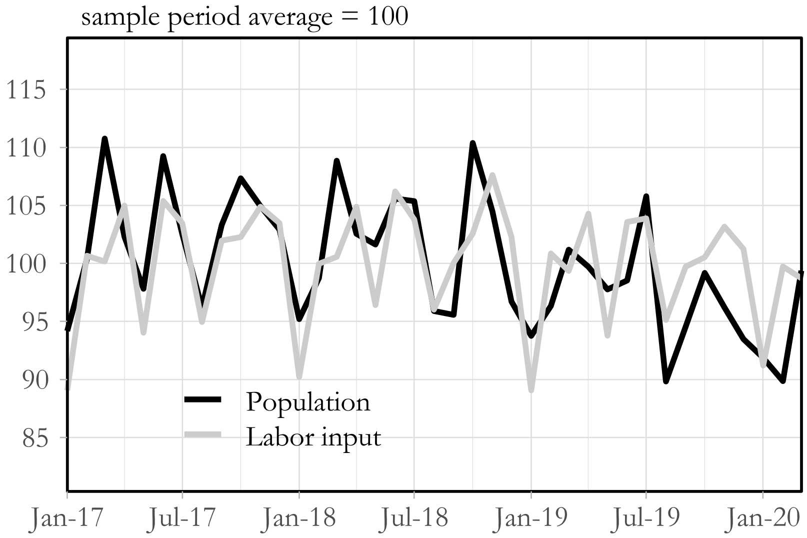


Figure 8: Population and labor input in the manufacturing industry

Sources: Agoop; Ministry of Health, Labour and Welfare.

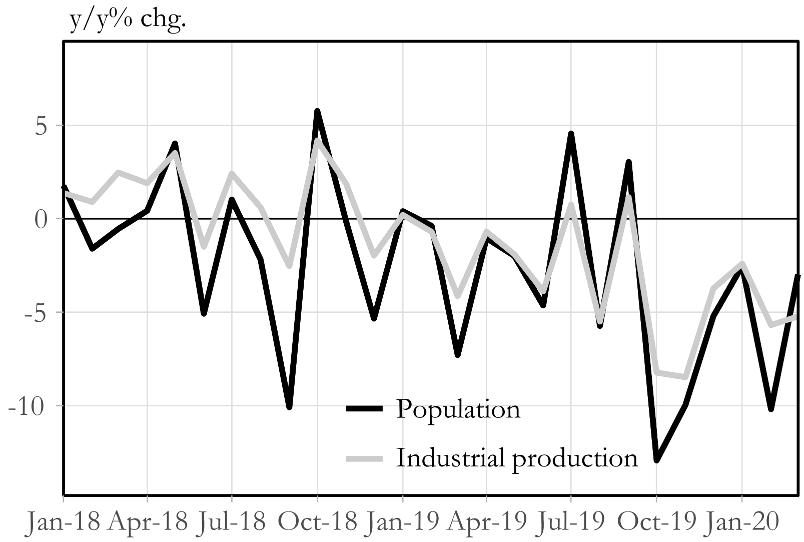


Figure 9: Indicator based on the hourly population for industrial production

Sources: Agoop; Ministry of Economy, Trade and Industry.

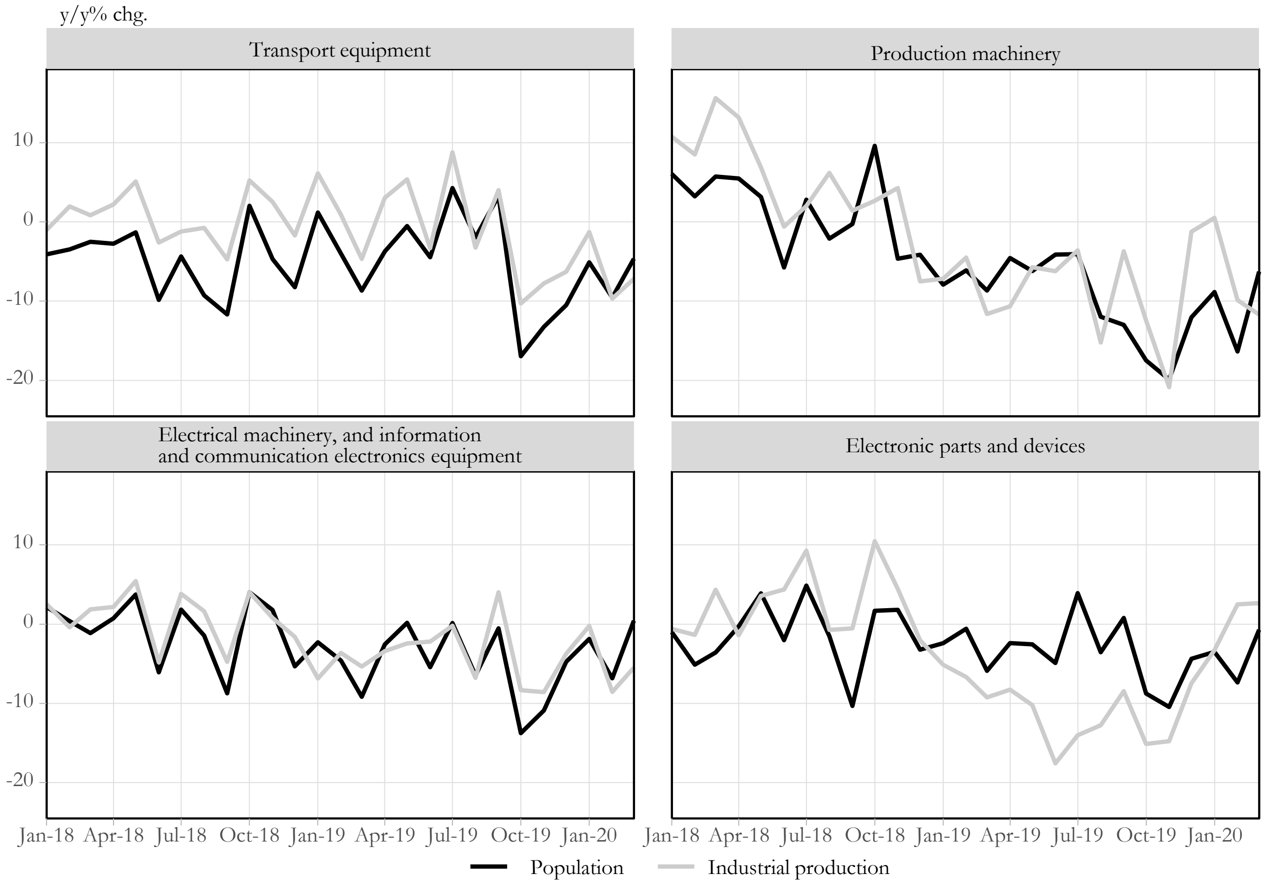


Figure 10: Production indicator based on the hourly population by industry

Sources: Agoop; Ministry of Economy, Trade and Industry.

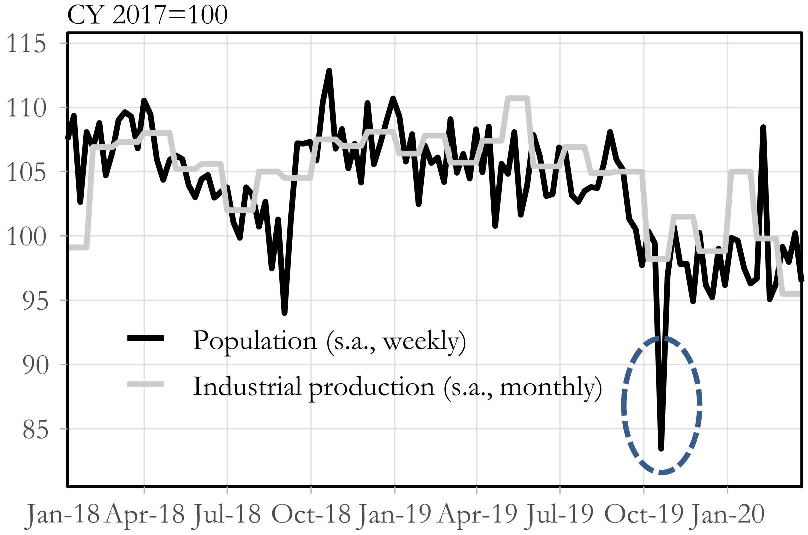


Figure 11: Weekly production indicator (transportation equipment)

Sources: Agoop; Ministry of Economy, Trade and Industry.

Table 1: Correlations between the EIG and consumption indexes for the food service industry

	Before clustering		Cluster 1	
	up to Feb. 2020	up to Mar. 2020	up to Feb. 2020	up to Mar. 2020
Food services	0.59	0.85	0.66	0.86
Bars and pubs	0.47	0.75	0.51	0.77

Notes: The table shows the correlations between the economic indicator from GPS data (EIG) for the food service industry and the corresponding indexes from JCB Consumption NOW based on credit card transaction data.
Sources: Agoop; Gurunavi; Ministry of Land, Infrastructure, Transport and Tourism; NOWCAST, Inc./ JCB, Co., Ltd., "JCB Consumption NOW." .

Table 2: Correlations between labor input and production

Industry	Correlation
Production machinery	0.75
Transportation equipment	0.72
Fabricated metal products	0.43
Plastic products	0.23
Pulp, paper and paper products	0.18
Ceramic, stone and clay products	0.15
Electronic parts, devices and electronic circuits	0.12
Food, beverages, and tobacco	0.05

Note: The table shows the correlation based on year-over-year changes in the period from January 2014 to December 2019. Labor input = Total hours worked $\times$ Regular employees (from the Monthly Labour Survey).
Sources: Ministry of Economy, Trade and Industry; Ministry of Health, Labour and Welfare.

Table 3: Criteria for Sunday and daytime ratios, and average population

Industry	Sunday ratio upper threshold	Day time ratio lower threshold	Average population lower threshold	Correlation
Transportation equipment	0.5	0.6	10	0.86
Production machinery	0.1	2	40	0.69
Information and communication electronics equipment	0.9	1	80	0.68
General-purpose machinery	0.5	1.8	40	0.63
Food, beverages and tobacco	0.1	1	60	0.60
Business oriented machinery	0.5	2	10	0.54
Non-ferrous metals and products	0.1	1.4	20	0.49
Electrical machinery, equipment and supplies	0.1	0.8	10	0.42
Electronic parts, devices and electronic circuits	0.9	1.4	10	0.31
Chemical and allied products	0.9	0.6	10	0.20

$Note$: Labor input = Total hours worked $\times$ Regular employees (from Monthly Labour Survey). Each ratio is calculated using mobility data on the weeks of which the previous and following weeks have five business days as well as those weeks themselves in order to eliminate the effect of public holidays. Sunday ratio = average population on Sunday / average population on weekdays. Daytime ratio = average population from 9:00 a.m. to 4:59 p.m. / average population from midnight to 4:59 a.m.
$Sources$: Agoop; Ministry of Health, Labour and Welfare.

Table 4: Criterion for holiday inclusion and hours

Industry	Conditions			Correlation
	Holidays included	Start hour	End hour	
Transportation equipment	yes	17	18	0.85
Production machinery	yes	18	21	0.82
Iron, steel and Non-ferrous metals	yes	8	12	0.78
Electrical machinery, and information and communication electronics equipment	no	16	17	0.72
Fabricated metals	no	16	19	0.71
Plastic products	yes	17	18	0.69
General-purpose and business oriented machinery	yes	10	11	0.65
Pulp, paper and paper products	no	9	10	0.61
Ceramics, stone and clay products	yes	14	15	0.55
Electronic parts and devices	yes	8	9	0.36
Chemicals	no	16	17	0.35
Foods and tobacco	no	22	23	0.30

$Note$: "Start hour" and "end hour" indicate the start and end time of the selected sample window, respectively. For example, the row with a "start hour" of 18 and "end hour" of 21 means the data taken from 18:00 to 21:59.
$Sources$: Agoop; Ministry of Health, Labour and Welfare.

Nowcasting Economic Activity with Mobility Data*

Kohei Matsumura (Bank of Japan), Yusuke Oh (BoJ), Tomohiro Sugo (BoJ), and Koji Takahashi (BIS)

IFC-Bank of Italy workshop

February, 2022

* The views expressed here are those of authors and do not necessarily reflect those of the Bank of Japan or BIS.

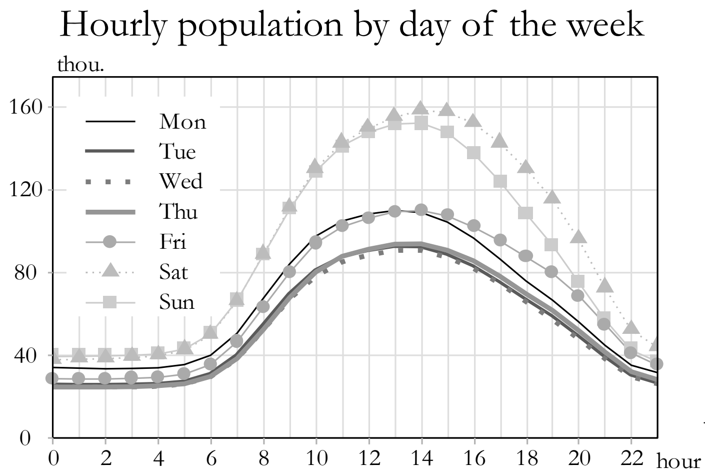
Data

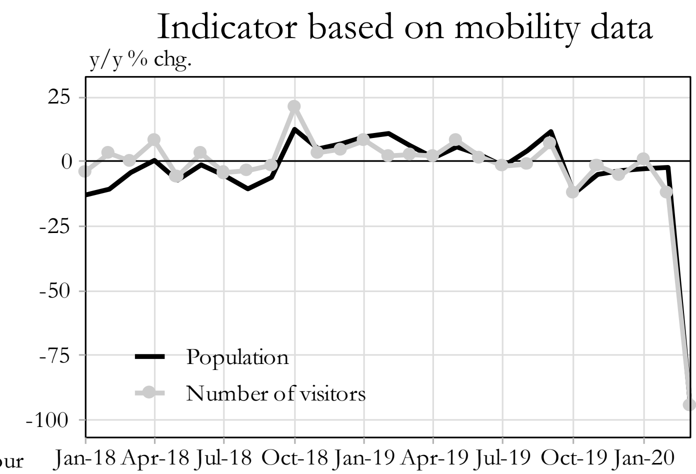
- ➢Collected by smart phone apps in Japan.
- ➢The number of people who stayed in a mesh
- ➢A mesh is defined as **100m × 100m** square.
  - ✓20 mil. Meshes in total.
- ➢Hourly data from Jan. 2017 to Mar. 2020.
- ➢Combine with other POI data
  - Economic Census
    - ✓ Factory addresses, number of employees, sales, etc.
  - National Land Numerical Information
    - ✓ Coordinates of stations and airports, and use of lands.
  - Information provided by private companies
    - ✓ Coordinates of facilities of interest (e.g. restaurants) based on their names and addresses.

Mesh ID	Year	month	day	hour	population
XX	2017	1	1	0	1
XX	2017	1	1	1	2
XX	2017	1	1	2	30
⋮	⋮	⋮	⋮	⋮	⋮
34 bil. records in total					
⋮	⋮	⋮	⋮	⋮	⋮
ZZ	2020	3	31	20	1
ZZ	2020	3	31	23	10

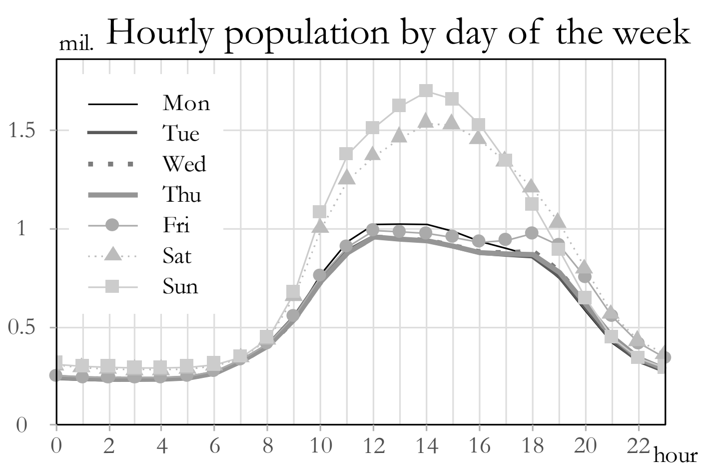
Amusement parks and shopping malls

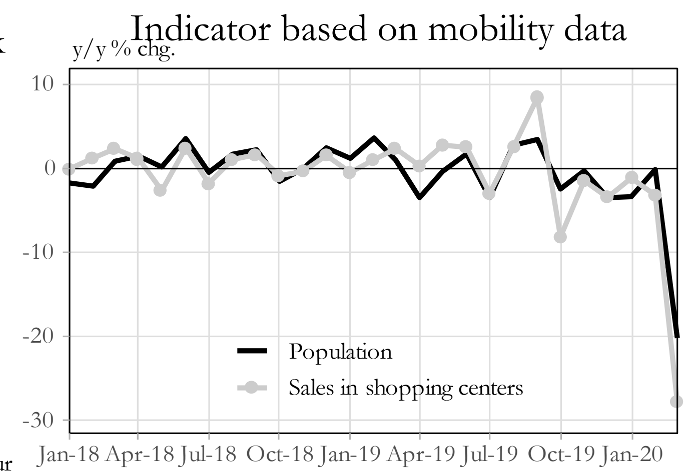
Amusement parks





Shopping malls



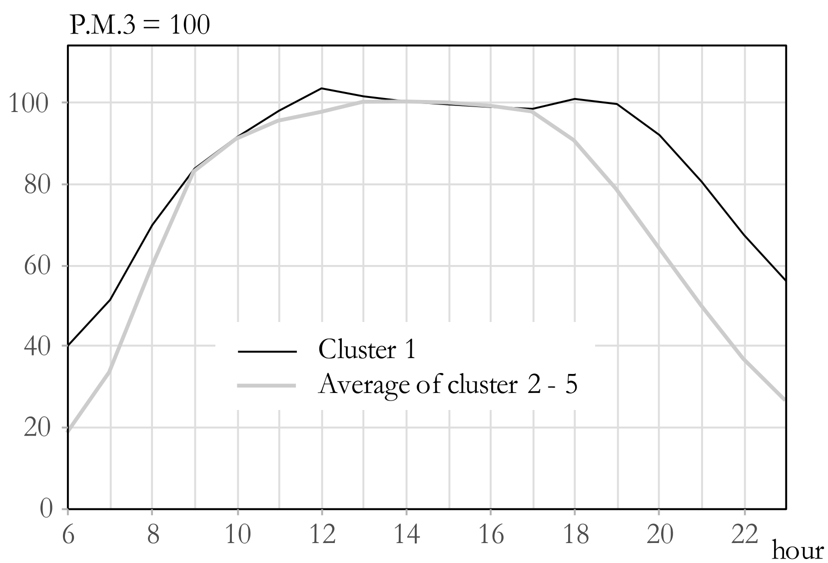


Restaurants

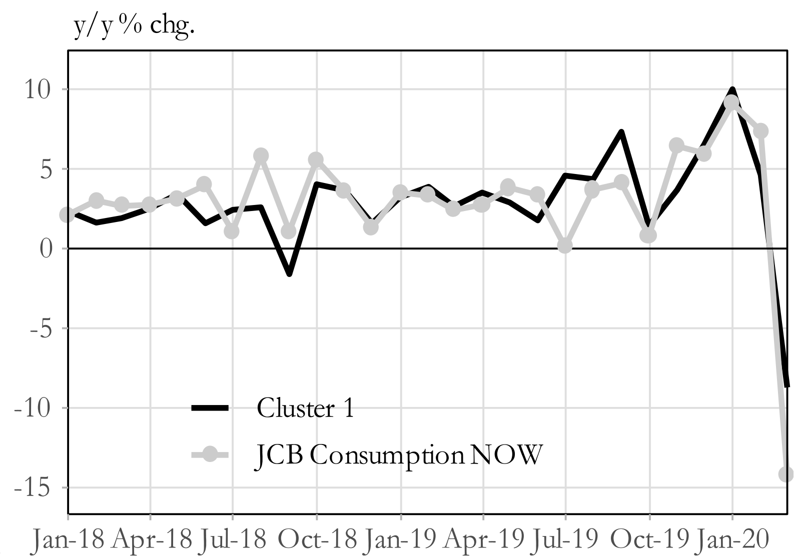
- Small size for each one but the number of restaurants is enormous.
- Located in areas where many various different activities are conducted.

1. Focus on non-residential areas based on the ratio of population in daytime
2. Use Grunavi (restaurant guide service) API and extract meshes that include more than 300 restaurants
3. Cluster the selected meshes into 5 groups, using k-medoids

Hourly population on weekdays by cluster



Indicator based on the hourly population

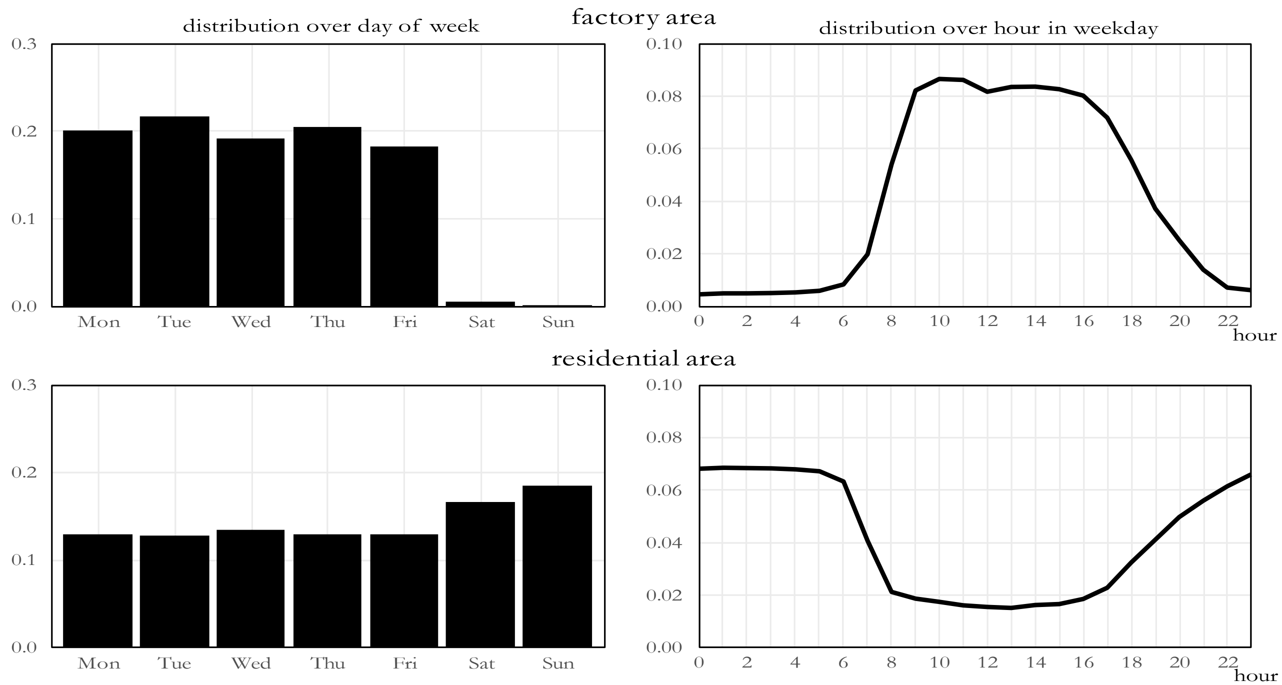


Sources: Agoop; Gurunavi; Ministry of Land, Infrastructure, Transport and Tourism; NOWCAST, Inc./ JCB, Co., Ltd., "JCB Consumption NOW."

Production: Identifying meshes representing factories

1. Among factories listed in Economic Census for Business Activity whose area are above 10,000 m^2, select top 10,000 factories in terms of value-added.
2. We collect candidate meshes around the registered address by using information on its area.
3. Focus on several features (Sunday ratio, etc.) to remove unrelated ones.

For each industry, grid search three thresholds (day time ratio, Sunday ratio, and average population) such that the correlation between population and labor input (from official statistics) is highest.

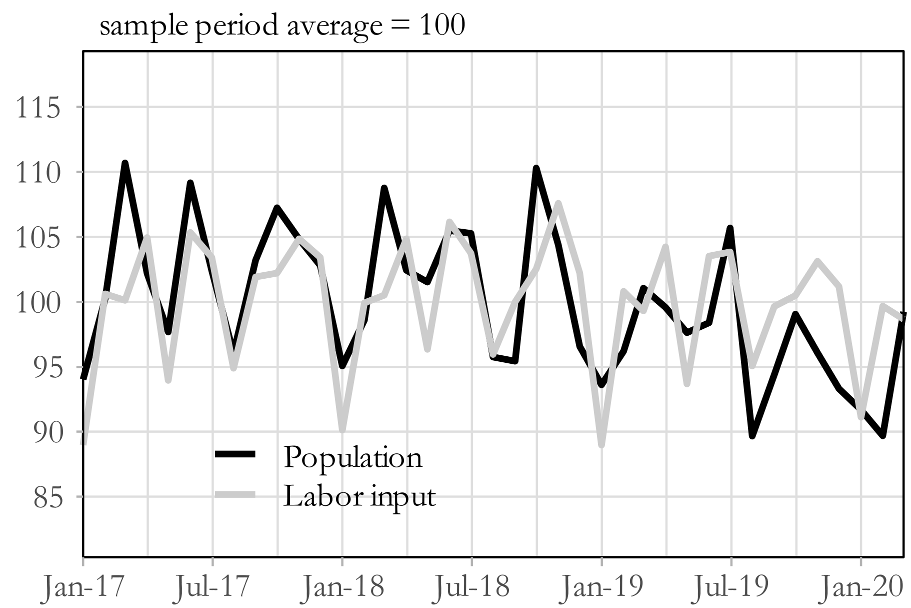


Mobility data and labor input

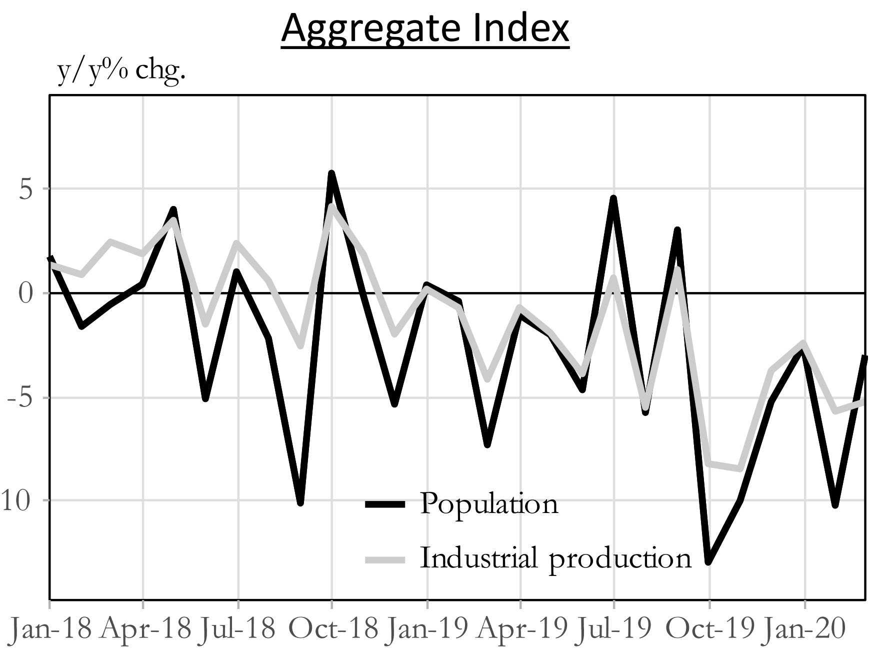
➢Aggregating population over all hours and days does not necessarily approximate labor inputs well.

e.g. fluctuations can be driven by facility maintenance workers.

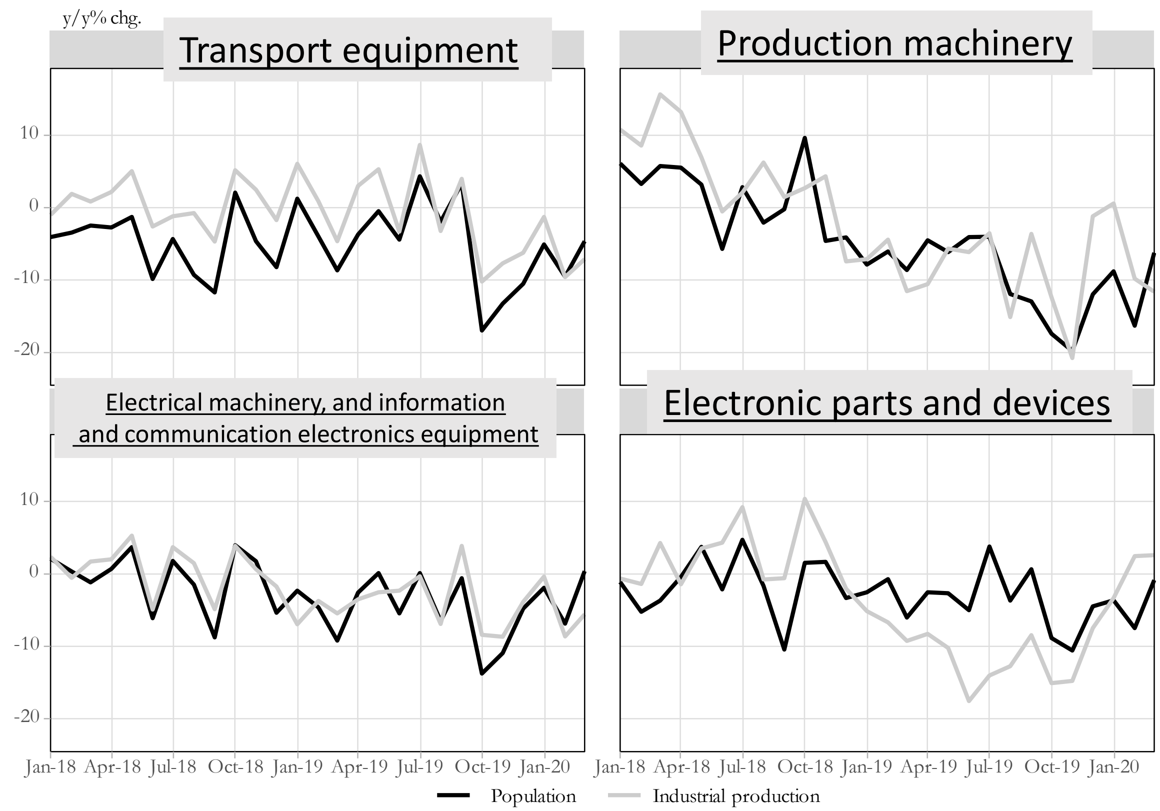
➢Grid search three parameters (start and end time, and whether including weekends) to determine the best time window.



Mobility data and industrial production



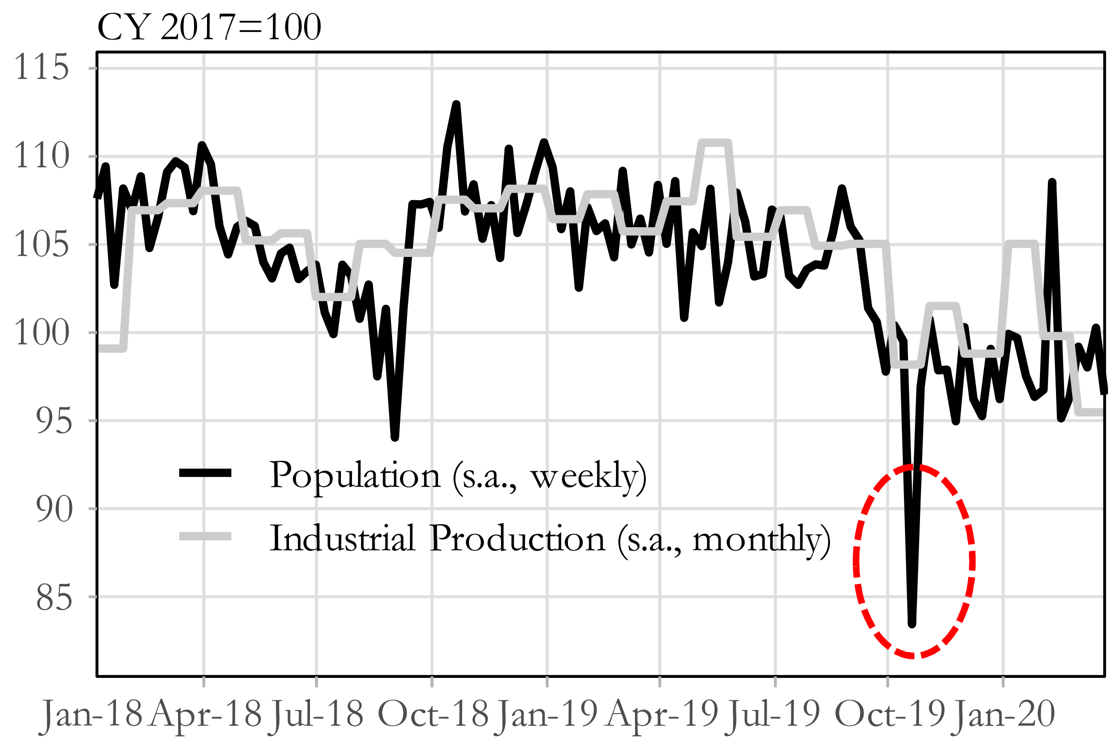
Sources: Agoop; Ministry of Economy, Trade and Industry.



Extension: High-frequency Analysis

- ➢ We construct weekly index by aggregating our indices at the weekly level and removing seasonal effects.
  - ✓ The index capture the impact of Typhoon Hagibis on production in late Oct. 2019.

Weekly production indicator (transportation equipment)



Sources: Agoop; Ministry of Economy, Trade and Industry.