Bank of Japan Working Paper Series

A Nowcasting Model of Industrial Production using Alternative Data and Machine Learning Approaches

Kakuho Furukawa*
kakuho.furukawa@boj.or.jp

Ryohei Hisano**
hisanor@g.ecc.u-tokyo.ac.jp

Yukio Minoura***
yukio.minoura@boj.or.jp

Tomoyuki Yagi*
tomoyuki.yagi@boj.or.jp

No.22-E-16
Nov. 2022

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

* Research and Statistics Department
** The University of Tokyo
*** Research and Statistics Department (currently Financial System and Banking Examination Department)

A Nowcasting Model of Industrial Production using Alternative Data and Machine Learning Approaches*

Kakuho Furukawa† Ryohei Hisano‡ Yukio Minoura§ Tomoyuki Yagi**

November 2022

Abstract

Recent years have seen a growing trend to utilize "alternative data" in addition to traditional statistical data in order to understand and assess economic conditions in real time. In this paper, we construct a nowcasting model for the *Indices of Industrial Production* (IIP), which measure production activity in the manufacturing sector in Japan. The model has the following characteristics: First, it uses alternative data (mobility data and electricity demand data) that is available in real-time and can nowcast the IIP one to two months before their official release. Second, the model employs machine learning techniques to improve the nowcasting accuracy by endogenously changing the mixing ratio of nowcast values based on traditional economic statistics (the *Indices of Industrial Production Forecast*) and nowcast values based on alternative data, depending on the economic situation. The estimation results show that by applying machine learning techniques to alternative data, production activity can be nowcasted with high accuracy, including when it went through large fluctuations during the spread of the COVID-19 pandemic.

JEL Classification: C49, C55, E23, E27
Keywords: Industrial production, Mobility data, Electricity data, Nowcasting, Machine learning, COVID-19

* The authors thank Kosuke Aoki, Ryo Jinnai, Seisaku Kameda, Jouchi Nakajima, and staff members of the Bank of Japan for their valuable comments. All remaining errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.
† Research and Statistics Department, Bank of Japan (kakuho.furukawa@boj.or.jp)
‡ The University of Tokyo (hisanor@g.ecc.u-tokyo.ac.jp)
§ Research and Statistics Department, Bank of Japan (currently Financial System and Bank Examination Department, yukio.minoura@boj.or.jp)
** Research and Statistics Department, Bank of Japan (tomoyuki.yagi@boj.or.jp)

1. Introduction

Recent years have seen a growing trend – both in Japan and abroad – to utilize "alternative data" in order to understand and assess economic conditions in real time.[1,2] While the speed and amplitude of economic fluctuations have increased particularly during the COVID-19 pandemic, traditional economic statistics have a time lag of several weeks to months before they are released. As a result, the use of alternative data that becomes available in real-time is increasing.[3]

For instance, in the United States, the Federal Reserve Bank of New York combines traditional economic statistics (e.g., initial claims for unemployment insurance) and alternative data (e.g., electricity output) to nowcast the current state of the economy on a weekly basis (Lewis et al. (2020)). In Japan, Nakazawa (2022) and Okubo et al. (2022) construct nowcasting models for GDP and personal consumption respectively using alternative data. These studies show that the use of alternative data is useful for nowcasting. Furthermore, a key feature of recent nowcasting literature is that an increasing number use machine learning models in addition to the econometric methods traditionally used in economic analyses (such as linear regression analysis).[4] Fornaro (2020), for example, shows that machine learning models can improve the accuracy of nowcasting of production activity.

In this paper, we construct a model for nowcasting production in the manufacturing sector – one of the most important factors when assessing economic conditions – using alternative data and machine learning approaches. While the Ministry of Economy, Trade and Industry (METI) releases the *Indices of Industrial Production* (IIP) as a gauge of production trends in Japan, the figures for April 2022, for instance, were published only at the end of May, i.e., with a time lag of about one month. The aim of the model developed in this paper is to assess, with the help of alternative data, actual production activity more than one month before the release of the official statistics.

---

[1] Following Kameda (2022), we collectively refer to data other than traditional economic data (e.g., monthly and quarterly macroeconomic indicators and financial disclosures of listed companies) as alternative data.

[2] For instance, BIS (2021) surveys the use of alternative data by central banks and finds that its use has grown rapidly in recent years.

[3] Taking quarterly GDP statistics as an example, in the United States and the euro area, these are published at the end of the month following the relevant quarter, while in Japan they are published in the middle of the second month following the relevant quarter.

[4] Although there is no established definition of "machine learning," in this paper we refer to machine learning models as models that use methods such as random forests and gradient boosting (see below for details) that differ from those traditionally used in economic analyses.

Previous studies using alternative data to analyze production in the manufacturing sector include Suimon and Yanai (2020) and Matsumura et al. (2021). These studies use GPS location data from mobile phones to measure the number of people located in factory areas and show that this is highly correlated with production activity. Moreover, Suimon et al. (2019) argue that industrial production can be nowcasted using electricity demand, while Kawamura et al. (2021) nowcast the IIP using trucks' car navigation data. However, Suimon and Yanai (2020) and Matsumura et al. (2021) only show correlations between alternative data and the IIP and do not develop models that nowcast the IIP. Meanwhile, Suimon et al. (2019) and Kawamura et al. (2021) do not analyze production activity by industry.

In the nowcasting model in this paper, we follow Matsumura et al. (2021) and Suimon et al. (2019) and use alternative data on the number of persons located in particular factory areas and data on electricity demand. We also use information from traditional economic statistics, the *Indices of Industrial Production Forecast* (hereafter, IIP Forecast). Building a model that uses alternative data in addition to traditional economic statistics should make it possible to follow developments in production even when the economy is hit by large shocks and it is difficult to capture developments in production before the release of the IIP using only traditional economic statistics. Furthermore, we aim to improve the accuracy of the model by utilizing machine learning techniques.

This paper is organized as follows. Section 2 describes the data we use, while Sections 3 and 4 outline the construction and results of the model, respectively. Section 5 provides a summary and concludes.

2. Data

2.1. Alternative Data

To consider production activity in the manufacturing sector, we focus on labor input and capital input. The latter can be explained in terms of the capital stock and the capital utilization rate (operating ratio). While the capital stock tends to fluctuate more in the medium to long term reflecting corporate strategies, etc., short-term production fluctuations tend to be brought about by fluctuations in labor input and the capital utilization rate. Therefore, if we can find proxy variables for labor input and the capacity utilization rate in our nowcasting using alternative data, this should improve the accuracy of nowcasting short-term production fluctuations. Based on these considerations, we use mobility data as a proxy for labor input and electricity data as a proxy for the capital utilization rate.

Construction of the "Mobility Index" using Mobility Data

Following Matsumura et al. (2021), we use the mobility data provided by Agoop Inc. This data consists of hourly estimates for the number of persons in each 100 meter square mesh of Japan, and is derived from GPS location data obtained by the company with permission from its smartphone application users. Matsumura et al. (2021) capture production activity at each factory by identifying the mesh containing the factory and aggregating the number of persons in the mesh. We use the database created by Matsumura et al. (2021) to create the "mobility index" to measure production activity in the manufacturing sector.[5] The specific calculation method is as follows.

First, we standardize the number of persons at factory j from the first day to the dth day of month t.

$$Index_{j,t,d} = \frac{Number\ \ of\ \ persons_{j,t,d}}{\overline{Number\ \ of\ \ persons_j}} \times 100. \tag{1}$$

$Number\ \ of\ \ persons_{j,t,d}$ is the total number of persons in the mesh associated with factory j from the first day to the dth day of month t.[6,7] $\overline{Number\ \ of\ \ persons_j}$ is the average of the total number of persons at factory j per month between January 2017 and December 2019 (before the outbreak of COVID-19). The mobility data for a given day are available with a lag of a few days.

Second, we calculate the "mobility index" for industry i by weighting $Index_{j,t,d}$ for each factory using the value-added of goods produced at that factory:[8]

---

[5] Matsumura et al. (2021) use microdata of the 2016 *Economic Census of Business Activity* published by METI, which contain address information of factories in the manufacturing sector, to identify meshes that are related to factories. Therefore, it should be noted that our analysis does not take into account production activities of factories that were built after the census was conducted.

[6] When aggregating the data, Matsumura et al. (2021) used the one-hour time window that has the highest correlation with the year-on-year change in the IIP (for example, for the transportation equipment industry, they use the number of persons between 17:00 and 18:00). However, when we extend the data used by Matsumura et al. (2021) to September 2021 for this paper, we find that the one-hour time window with the highest correlation differs substantially depending on the period. For this reason, we do not specify the time window but use the number of persons during the entire 24 hour period.

[7] Specifically, we aggregate the average hourly number of persons between 0:00h on the first day of month t and 24:00h on the dth day of the same month.

[8] The value added of goods produced at each factory is obtained from microdata from the 2016 *Economic Census of Business Activity* published by METI.

$$Mobility\ \ Index_{i,t,d} = \sum_{j \in S_i} \frac{VA_j}{TotalVA_i} \cdot Index_{j,t,d}, \tag{2}$$

where S_i is the set of factories classified into industry i, VA_j is the value-added of goods produced at factory j, and $TotalVA_i$ is the sum of the value-added of goods in all factories classified into industry i.

We calculate the mobility index for January 2017 through September 2021. The correlation coefficient between the mobility index and the IIP is 0.81, suggesting that the index is useful for nowcasting production activity (Figure 1(a)). The index is able to capture the decline in production in May 2020, when the impact of the spread of COVID-19 was particularly large. On the other hand, during the recovery in production in the first half of 2021, the year-on-year growth of the index remains below that of the IIP. This may be because the number of workers coming to work at factories was lower than before the pandemic even as production recovered, as work-from-home became more widespread in the manufacturing sector due to the prolonged restrictions on mobility caused by the pandemic.[9]

Comparing the correlation coefficients between the mobility index and IIP at the industry level shows that this is high for the transport equipment industry, while it is low for the electronic parts and devices industry (Figure 1(b), Table 1). As pointed out by Matsumura et al. (2021), the electronic parts and devices industry is highly capital-intensive, and our result suggests that nowcasting production activity using only the mobility index, which serves as a proxy for labor input, for highly capital-intensive industries is difficult. This is the reason that we also use the "electricity index" as a proxy for capital utilization, which we discuss next.

Construction of the "Electricity Index" using Electricity Demand Data

To construct the electricity index, we use data on electricity demand provided by the Organization for Cross-regional Coordination of Transmission Operators. Specifically, we employ data on the volume of the electricity demand by region (for 10 regions nationwide) and by time of day (hourly). The data is available free of charge on the organization's website on the following day.

[9] Calculating labor input (total actual hours worked, including work-from-home hours, multiplied by the number of employees) using the *Monthly Labor Survey* shows that it generally moves in a similar manner as the IIP during this period (Figure 2). It can be inferred that there may have been a shift to work-from-home for jobs that require less attendance at the factory (such as jobs in administrative departments rather than on the production line).

Since the demand for electricity at factories should increase as production activity increases, electricity data is likely to be useful for nowcasting production activity in the manufacturing sector. It should be noted, however, that the published electricity data does not provide information on the users, and that it contains electricity demand not only by the manufacturing sector but also by general households and the non-manufacturing sector (such as by commercial facilities) and is subject to heating and cooling demand depending on weather conditions (Figure 3). This means that it is not a straightforward matter to extract only the electricity demand attributable to manufacturing production activities. Against this background, we use electricity demand excluding weather factors to construct our electricity index and use it for nowcasting manufacturing production activities.[10]

$$Electricity\ \ Index_{t,d} = \sum_{1 \le j \le d} \sum_{i} exp\left[log\left(Elec_{i,t,j}\right) - Weather_{i,t,j}\right] \tag{3}$$

where $Elec_{i,t,j}$ is the daily electricity demand on day j in month t in region i (10 regions nationwide). $Weather_{i,t,j}$ is the estimated electricity demand attributable to weather factors, which we calculate by regressing the linear and quadratic terms of temperature and precipitation and the cross terms of the two for each region on the logarithm of the actual electricity demand.[11] As mentioned above, the electricity data does not contain industry-level information, so that the index created here is the same for all industries.

As with the mobility index, we calculate the electricity index for the period from January 2017 to September 2021. The correlation coefficient between the index and the IIP is high at 0.73. Although it is lower than the correlation coefficient between the mobility index and the IIP (0.81), it suggests that the index is useful for nowcasting production activity (Figure 4). The electricity index captures movements in the IIP relatively well, especially during the recovery in production in the first half of 2021, when they were not fully captured by the mobility index. As mentioned above, while the number of factory workers was likely still below pre-pandemic levels during this phase,

---

[10] Here, we assume that electricity demand in the manufacturing sector fluctuates mainly due to production factors and that the electricity index excluding weather factors captures production trends. While we assume that, in the short run, electricity demand by households and the non-manufacturing sector fluctuates mainly due to weather factors, we cannot separate out fluctuations in demand by households and the non-manufacturing sector completely. It should therefore be noted that the electricity index in this paper is contaminated by demand trends outside the manufacturing sector.

[11] To gauge weather-related demand fluctuations, we tried various combinations of explanatory variables including temperature, humidity, and rainfall data, and selected the above combination, for which the correlation between the calculated electricity index and the IIP was the highest.

it seems possible to track production activity by looking at electricity use. This suggests that the mobility index and the electricity index capture production activities from different angles, and that constructing a nowcasting model using both of these indices together is likely to lead to improved accuracy.

Looking at the correlation coefficients between the electricity index and the industry-level IIP shows that they are particularly high in the iron, steel and non-ferrous metals industry and the transport equipment industry (Table 2).

2.2. Traditional Economic Statistics (*Indices of Industrial Production Forecast*)

In addition to the alternative data described above, we also use the IIP Forecast published by METI. This statistic is based on a survey of the production plans of major firms at the beginning of each month for the month and the following month in question (hereafter referred to as the "current month IIP Forecast " and the "following month IIP Forecast," respectively), and is published at the end of each month. Thus, the production forecast (i.e., firms' production plans) for April, for example, is released at the end of March (production plans as of the beginning of March) and at the end of April (production plans as of the beginning of April), which provides valuable data for assessing the future and current production activities in the manufacturing sector. However, as pointed out by METI (2020), due in part to differences in the sample firms, the IIP Forecast tends to be (1) higher and (2) more volatile than the IIP. We try to correct these biases and use the adjusted IIP Forecast values for our nowcasting model.[12] In the following analysis, unless otherwise noted, whenever we refer to the IIP Forecast, we are talking about the bias-adjusted values.

As shown in Table 3, the correlation coefficients between the industry-level IIP and the bias-adjusted IIP Forecast are high both for the current month IIP Forecast and the following month IIP Forecast (0.91 and 0.97 for the mining and manufacturing sector, respectively), confirming that we can predict firms' actual production activities with a high degree of accuracy using their production plans. However, the IIP Forecast is likely to fail to capture changes in production activity in the event of a major shock to the economy such as a natural disaster or a pandemic that occurs after the survey. Indeed, the IIP Forecast failed to capture the large drop in production in May 2020 due to the outbreak of the pandemic (Figure 5(a)).

[12] While *Statistics Survey of Current Industrial Production*, based on which the IIP is constructed, surveys around 14,000 firms about 412 goods, the IIP Forecast surveys around 800 firms about 186 goods.

To summarize the discussion on the three types of data used in this paper, the IIP Forecast – traditional economic statistics – is highly correlated with the IIP and can be regarded as a highly accurate predictor of production trends in the manufacturing sector. However, in the event of a major shock to the economy after the survey has been conducted, such as an outbreak of an infectious disease, the IIP Forecast is unable to capture changes in production activity and is likely to be less accurate. For example, this is the case when a shock to production occurs in April between the time of the IIP Forecast survey (the beginning of the month) and the end of the month.[13,14] In such cases, alternative data that are available on a daily basis (such as the mobility index and the electricity index) can be used to obtain information that cannot be captured by the IIP Forecast, providing a more accurate understanding of production trends. In the next section, we construct a nowcasting model based on the characteristics of these different data types.

3. Nowcasting Model

3.1. Nowcast Targets and Nowcast Points in Time

We build nowcasting models for the production of each of the nine manufacturing industries (at the major industry classification level, namely: iron, steel and non-ferrous metals; fabricated metals; production machinery; general-purpose and business oriented machinery; electronic parts and devices; electrical machinery and information and communication electronics equipment; transport equipment; chemicals; and pulp, paper and paper products) as well as industrial production overall. Therefore, we build 10 models in total.

Since the alternative data described in the previous section is available on a daily basis, we can construct nowcasting models on a daily basis. We construct a nowcasting model for the current month's IIP using information in the first week (day 7), the second week (day 14), the third week (day 21), and the end of the month, respectively (Table 4). As mentioned earlier, the data for the actual production in April 2022, for example, was released at the end of May. Hence, the nowcasting model we develop allows us to identify

[13] The IIP Forecast surveys production plans for the current month and the following month as of the first day of each month. The submission deadline for the survey is the 10th of each month.
[14] In addition, even if the shock itself occurs before the survey, if the external environment changes so rapidly that revisions to production plans have not kept pace, the IIP Forecast will likely be less accurate.

production trends about one to two months before the release of the actual production data.

3.2. Explanatory Variables

The dependent variable in the estimation is the relevant IIP, while the explanatory variables consist of (1) the mobility index, (2) the electricity index, and (3) the IIP Forecast. All variables are converted into year-on-year rates of change. For the IIP Forecast, we use the latest values available at the time the model is estimated – the following month IIP Forecast for the model for the first to third week, and the current month IIP Forecast for the model for the end of the month. We also try out adding one-period lags of the explanatory variables. Specifically, we consider several combinations of explanatory variables: (1) using only alternative data, (2) using the IIP Forecast in addition to the alternative data, and (3) using both the alternative data and the forecast indices and also adding lag terms (Table 5). When using the alternative data, we also consider specifications in which only one of the two indices is used. Overall, we test a total of nine specifications.

3.3. Estimation Models

Looking at previous studies, various types of models, ranging from traditional econometric models (such as ordinary least squares) to machine learning models, which have seen remarkable advances in recent years, have been used to nowcast economic indicators and/or forecast the future. However, there is no one-size-fits-all model that can consistently achieve a high prediction accuracy. The reason is that while simple models may fail to capture developments in the data well, overly complex models tend to adjust excessively to the existing data and make out-of-sample forecasts that are less accurate than those of simple models (Christian and Griffiths (2017)). This phenomenon is called "overfitting" and is particularly likely to occur when the sample size is small. Therefore, when constructing a nowcasting model, it is important to select a model with "moderate complexity" to avoid overfitting, while taking into account the sample size and characteristics of the data used. Based on these considerations, we try the following three types of models.

Linear regression models

Linear regression models are simple models that assume that the relationship between the independent and dependent variables is linear. Such models have the advantage of being less prone to overfitting. On the other hand, such models assume that the explanatory variables do not interact with each other. It has also been pointed out that if the number of explanatory variables becomes too large, overfitting may occur even in linear regression models, reducing the prediction accuracy (Stock and Watson (2006)).

Regression models using machine learning techniques

The next type of models is regression models using machine learning techniques. In this paper, we try two machine learning models that can analyze interactions among explanatory variables and nonlinearity: the random forests and gradient boosting approaches.[15] Both models can be estimated without making any a priori assumptions about the relationship between the independent and dependent variables and are widely used in the literature.[16] While technical summaries of both models can be found in, for example, Hastie et al. (2009), both models essentially "learn" by building a large number of decision trees. Random forest models are estimated by repeating the step of building decision trees based on sampling with replacement. In gradient boosting, each step involves building the next decision tree based on the decision trees built in the previous steps, so that the final prediction accuracy is higher. In other words, the model is built based on a succession of decision trees.

In general, machine learning models are more complex than linear regression models. When strong interactions among explanatory variables and/or nonlinearity are present, using a more complex model is likely to improve prediction accuracy. On the other hand, it should be noted that when the sample size is limited, as in the alternative data in our analysis, the model is more likely to suffer from overfitting, as discussed earlier.

A mixed model

As discussed in Section 2, in "normal times," – i.e., times when the economy is not subject to a major unexpected shock such as a natural disaster or a pandemic – production activity is likely to be forecasted with reasonable accuracy using only the IIP Forecast. In this case, the advantages of using alternative data – the ability to take into account

[15] Another advantage of the random forests and gradient boosting approaches is that the estimation results tend to be stable even when there are many explanatory variables. However, since the number of explanatory variables in this analysis is not very large, this advantage is unlikely to be substantial.

[16] Many previous studies show that the prediction accuracy of economic indicators can be improved by using random forests and gradient boosting (Fornaro (2020), Batarseh et al. (2020), Chapman and Desai (2021)). However, it should be noted that the benefits of using machine learning models in this analysis may not be as large as in previous studies since the sample size of the data used is smaller than in these previous studies.

information not included in the IIP Forecast – may be outweighed by the disadvantages, i.e., the possibility of overfitting due to a larger number of explanatory variables.

Based on these advantages and disadvantages, we propose here a "mixed model" as a unique method to obtain predictions that mixes different approaches. In machine learning analysis, the method of using multiple approaches together to obtain new predictions is called "blending" and is known to be useful for improving prediction accuracy (Timmermann (2006), Bolhuis and Rayner (2020)).

Specifically, we use the following formula to obtain predictions:

$$p \times Pred^{alt} + (1-p) \times IIP\ \ Forecast \tag{4}$$

where $Pred^{alt}$ is the predicted value obtained by linear regression models with alternative data. p is a variable parameter that varies with economic conditions and takes a value between 0 and 1. This value is calculated based on the movement of $Pred^{alt}$ and the IIP Forecast at a given time, using the two machine learning approaches (random forests and gradient boosting). In order to improve the prediction accuracy of equation (4), p is small in normal times when the variation of $Pred^{alt}$ is relatively small, since the prediction accuracy of the IIP Forecast itself is likely to be reasonably high, as mentioned earlier. On the other hand, when a large shock to the economy causes a relatively large change in $Pred^{alt}$, the informational value of the alternative data is expected to increase, and p is expected to become large.[17,18]

To summarize, we construct a total of 45 nowcasting models (five different models using nine variable combinations) for each of the 10 nowcast targets (industrial production overall as well as nine industries) and four forecast time points (three weekly time points and the end of the month) for comparison.[19,20,21]

[17] See Appendix 1 for details of the estimation method using the mixed model.

[18] This point was also made by Goshima et al. (2019), who argue that alternative data (high-frequency data) are superior to existing statistics when the economy is volatile.

[19] There are five types of models: a linear regression model, the regression models with random forests and gradient boosting, and the mixed models with random forests and gradient boosting.

[20] For some industries, such as the food and tobacco industry and the textiles industry, it is not possible to construct a nowcasting model because the IIP Forecast is not published for these industries.

[21] There is no lead/lag relationship between the explanatory variables (the mobility index and the electricity index) and the dependent variable (the IIP) and the correlation is highest in the same period. Therefore, in constructing the model, we use the contemporaneous explanatory and dependent variables.

4. Results

4.1. Evaluation of Nowcast Accuracy

We build each of the models discussed in the previous section using data from January 2018 through September 2021. We then select the nowcasting model with the highest out-of-sample nowcasting accuracy for January 2020 through September 2021 as the best model.

While there are a range of measures for evaluating prediction accuracy, the one we employ is the root mean squared error (RMSE), which is widely used in the literature. Specifically, for instance, we first build a model using historical data through December 2019 and then perform nowcasting of the IIP for January 2020. We next perform nowcasting for February 2020 using historical data through January. We repeat this process until the nowcasting for September 2021 and select the model with the smallest RMSE as the best model. In the following, we use the IIP Forecast as a benchmark and focus on whether the nowcast accuracy of the calculated best model is better than the benchmark.

4.2. Estimation Results[22]

We start by reviewing the results of the nowcasting models for the IIP overall: in the first week (day 7), the benchmark has greater accuracy than the estimated models (Figure 6). At this point of the month, it appears that the limited availability of daily alternative data does not improve the accuracy of the model. From the second week onward, the nowcast accuracy of the best models exceeds that of the benchmark. At the end of the month, the best model is the one using machine learning techniques (the mixed model). The nowcasting accuracy improves with each passing week, implying that the use of alternative data leads to improved nowcasting accuracy by adding information on production activity during periods not covered by the IIP Forecast. Looking at the estimation results of the nowcasting model as of the end of the month, it can be seen that the best model may be more useful than the benchmark especially in times of sudden changes in production. For instance, in May 2020, when production dropped sharply due to the spread of the pandemic, the benchmark was not able to sufficiently predict the drop

[22] In assessing the state of the economy, policy makers often consider month-on-month and/or quarter-on-quarter growth rates using seasonally adjusted series. Since the time series of the alternative data used in this paper are short and it is therefore difficult to perform seasonal adjustments, we build the models on a year-on-year basis. Appendix Figure 1 shows the nowcasting results on a year-on-year basis converted to a month-on-month basis, and as in the case of the analysis on a year-on-year basis in this section, the RMSE of the best model is lower than the RMSE of the IIP Forecast.

in production while the best model was able to do so with high accuracy at the end of the month.

Next, looking at the estimation results of the nowcasting model at the industry level, the nowcast accuracies of the best models exceed those of the benchmark in all major industries as of the end of the month (Figure 7). In particular, the nowcast accuracy improves significantly for industries for which the benchmark accuracy is relatively low, such as the electronic parts and devices industry. In terms of the timing of nowcasting, the best models achieve higher nowcasting accuracy than the benchmark at a relatively early point in time for industries such as the electrical machinery and information and communication electronics equipment industry and the electronic parts and devices industry.

If we compare the average RMSE of the best models for the manufacturing sector overall and the nine industries at each point of the month with the benchmarks, the best models outperform the benchmarks from the first week on average and their comparative advantage becomes larger over each passing week (Figure 8).[23]

4.3. Discussion of Model Selection

Table 6 lists the best models that are selected for each industry and point of the month. Although the linear regression model, the simplest model, is selected in some cases, the mixed models are chosen in most cases. In this subsection we offer a few thoughts on the model selection based on a comparison of the RMSE of the different models (Figure 9).[24] In addition, we also consider the advantages of using alternative data.

First, comparing the RMSE of each model, the following points can be noted. First, the linear regression model in all specifications has smaller RMSEs than the random forests and gradient boosting machine learning regression models. Due to the limited sample size in this analysis, it is highly likely that overfitting occurred due to the complexity of the machine learning models. That said, the accuracy of the machine learning models is likely to improve in the future as more observations of the alternative data employed in this paper are accumulated and longer time series become available.

---

[23] Kawamura et al. (2021), who also attempt to nowcast the IIP on a weekly basis, find that the correlation coefficient between the IIP and the predicted value increases toward the end of the month.

[24] Machine learning models have complex algorithms and it is difficult to know how the predictions are calculated. Consequently, they are sometimes referred to as black boxes in contrast to white boxes (such as linear regression models) that can be easily interpreted. In this section, we discuss the model selection based on summary statistics.

Second, the mixed models have smaller RMSEs than the linear regression model and the machine learning regression models in all specifications. As discussed in the previous section, when nowcasting economic indicators using alternative data, linear regression models may not capture production activity well because they are too simple. On the other hand, using regression models based on random forests or gradient boosting may lead to overfitting when the sample is small. The estimation results in this paper suggest that employing mixed models such as the ones developed in this paper is a useful approach in such situations.

Furthermore, when comparing the nine different combinations of explanatory variables, the RMSE of Specification 9, which includes both the mobility index and the electricity index as explanatory variable, is the smallest. This implies that constructing models with multiple alternative data sets can help to capture various developments in production activity and improve the accuracy of nowcasting.

Next, we examine the benefits of using alternative data. To do so, in Figure 10 we calculate the relationship between the "realization rate" of the IIP Forecast, which is the IIP Forecast divided by the IIP (actual) and can be calculated ex post, and the value of p in equation (4), which is the weight assigned to the forecasted value based on the alternative data that is used for nowcasting. Looking at the relationship between the two, p, which is estimated from the variation in the alternative data, turns out to be smaller when the nowcast accuracy of the forecast index is high (around 100%) and larger when the accuracy is considerably different from 100%. In other words, the mixed models achieve higher prediction accuracy by attaching greater importance to information from the alternative data (i.e., assigning a larger value to p) during phases when the external environment changes substantially.[25]

5. Conclusion

In this paper, we constructed a nowcasting model for the IIP that combines traditional economic statistics (the IIP Forecast) and alternative data (mobility data and electricity data) and that employs machine learning techniques. The best model selected has greater nowcasting accuracy than the IIP Forecast, and the model allows more timely and

[25] In other words, the relationship shown in Figure 10 between the p estimated by the model and the "realization rate" of the IIP Forecast calculated ex post suggests that we can successfully predict the likelihood that the production plans indicated by the IIP Forecast will be revised by estimating a mixed model using changes in the alternative data.

accurate nowcasting of the IIP. The model has especially good nowcasting accuracy for the period when production activity changed rapidly during the COVID-19 pandemic.

A number of points are worth noting with regard to the analysis in this paper. First, while the alternative data used in this study has the advantage that it is available in a more timely manner than the IIP, it is not being compiled for the purpose of capturing production activities in the first place. This means that it may capture something other than production activities. For instance, the mobility index may remain high even when production is suspended if the number of persons at factories remains high due to reasons that are not directly related to production such as the large-scale inspection and repair of facilities or the construction of additional production lines. In such cases, the relationship between production and the mobility index may change, and the accuracy of the nowcasting model in this paper may be reduced. It should also be noted that the mobility index in this paper relies on business establishment data from the 2016 *Economic Census for Business Activity* and therefore does not take the establishment of new factories or relocation of existing factories into account.

The second point worth noting is that the time series of the alternative data used in this paper are rather short and the analysis is based on a limited sample. It is possible that the performance of the best model constructed in this paper changes in the future as the impact of the pandemic eases and the economy enters a new phase. It is also possible that the performance of the model could be further improved by using alternative data not used in this paper.[26]

In light of the above, in order to further improve nowcasting accuracy, it is necessary to accumulate alternative data while deepening our understanding of its characteristics and exploring optimal models while taking structural changes in the economy into account.

[26] Against this background, Appendix 2 presents an analysis using car traffic data as an example of an extension of the nowcasting model.

References

Batarseh, F. A., M. Gopinath, and A. Monken (2020), "Artificial Intelligence Methods for Evaluating Global Trade Flows," International Finance Discussion Papers 1296, Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/IFDP.2020.1296.

BIS (2021), "Use of Big Data Sources and Applications at Central **B**anks," IFC Report No.13 Feb. 2021.

Bolhuis, M.A. and B. Rayner (2020), "Deus ex Machina? A Framework for Macro Forecasting with Machine Learning," IMF Working Paper Series, No. 20/45.

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

Christian, B. and T. Griffiths (2017), *Algorithms To Live By: The Computer Science of Human Decisions*, William Collins.

Fornaro, P. (2020), "Nowcasting Industrial Production Using Unconventional Data Sources," ETLA Working Papers, No. 80

Goshima, K., H. Takahashi, and T. Yamada (2019), "Constructing Business Sentiment News Index Using Natural Language Processing and its Application to Predicting Volatility," *Monetary and Economic Studies*, Vol. 38(3), pp. 1-41 (available in Japanese only).

Hastie, T., R. Tibshirani, and J. H. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition*, Springer.

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

Kawamura, T., T. Kubo, and E. Yodoya (2021), "Predicting the Indices of Industrial Pdoruction Using Truck Car Navigation Data – An Attempt Using Machine Learning Approaches –," ESRI Discussion Paper Series (available in Japanese only).

Lewis, D., K. Mertens, J.H. Stock, and M. Trivedi (2020), "High Frequency Data and a

Weekly Economic Index during the Pandemic," Federal Reserve Bank of New York Staff Reports, No. 954.

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

METI (2020), "Can We Follow Production Activities with Production Plans? How to Make Use of the Indices of Industrial Production Forecast," METI Journal Online, https://meti-journal.jp/p/9255/ (available in Japanese only).

Nakazawa, T. (2022), "Constructing GDP Nowcasting Models Using Alternative Data," Bank of Japan Working Paper Series, No. 22-E-9.

Okubo, T., K. Takahashi, H. Inatsugu, and M. Takahashi (2022), "Development of 'Alternative Data Consumption Index': Nowcasting Private Consumption Using Alternative Data," Bank of Japan Working Paper Series, No. 22-E-8.

Stock, J. H., and M. W. Watson (2016), "Forecasting With Many Predictors," *Handbook of Economic Forecasting*, Vol. 1, pp. 515–554.

Suimon Y., K. Izumi, H. Sakaji, T. Shimada, and H. Matsushima (2019), "Real-time Economic Indicators of Manufacturing Industry based on High-frequency Electricity Demand Data," *Journal of the Japanese Society for Artificial Intelligence*, Vol 35(1), pp. 1-5 (available in Japanese only).

Suimon, Y. and M. Yanai (2020) "Using Mobile Phone GPS Data to Nowcast Auto Production Acitivty and Draw up an Equity Investment Strategy," JSAI Special Interest Group on Financial Informatics, pp. 9-12 (available in Japanese only).

Timmermann, A., (2006) "Forecast Combinations," in Elliott, G., C.W.J. Granger, and A. Timmermann, *Handbook of Economic Forecasting*, Elsevier.

Appendix 1. Estimation Method of the Mixed Model

The specific estimation approach for the mixed model used in this paper is as follows. The estimations are conducted for each nowcasting point in the month (week 1, week 2, week 3, and month end) and for each industry.

Step 1:	Estimate a linear regression model including alternative data as explanatory variables and calculate the predicted value of the IIP, $pred^{alt}$.[27]
Step 2:	Let z^{alt} and z^{pred} be the standardized values of $pred^{alt}$ and the corresponding IIP Forecast using the mean and standard deviation for each industry, respectively.
Step 3:	Compute variable y, which takes a value of 1 when the accuracy of the prediction of the linear regression model exceeds that of the IIP Forecast and 0 otherwise.
Step 4:	Train a model for which y is the dependent variable and z^{alt} and z^{pred} are the explanatory variables with random forests or gradient boosting.
Step 5:	Calculate the probability p that $y = 1$ based on the trained model.
Step 6:	Calculate the predicted value as follows: $$p \times pred^{alt} + (1 - p) \times IIP\ \ Forecast$$

[27] In the analysis, the same nine combinations of explanatory variables shown in Table 5 are tried in the linear regression model.

Appendix 2. Future Extensions of the Nowcasting Model

In this paper, we used mobility data and electricity data as proxies for inputs in production activities in constructing the nowcasting models for the IIP. The accuracy of the models could potentially be further improved through the use of a wider range of alternative data as they become available. Against this background, this appendix presents an analysis using car traffic volume data. Although this data is currently not available in a timely manner, the nowcasting model could be extended to include such data once it becomes available in real time.

Construction of the "Traffic Volume Index"

We use the traffic volume data provided by the Association for Promotion of Infrastructure Geospatial Information Disclosure (AIGID) to obtain the number of vehicles passing through approximately 40,000 observation points (general roads) nationwide.[28] This data is not published in real time – for example, January data becomes available in early March – so at present it is not very useful for a nowcasting model. However, the volume of traffic around a particular factory could serve as a proxy for information on inputs for production activities (such as the volume of raw materials and parts delivered).[29]

In order to utilize traffic volume data for nowcasting manufacturing production activities, we calculate the "traffic volume index" using the following steps. First, for each factory, observation points located within a 500 meter radius are identified. In some cases, there are multiple factories within a 500-meter radius of an observation point. We exclude these points from the analysis because they cannot be uniquely linked to a specific factory. Next, for each factory, the traffic volume at the identified observation points is aggregated to calculate the volume of traffic that is considered to have entered and exited the factory. Finally, using the same method as for the mobility index, we calculate the traffic volume index for each industry by taking the average of the traffic volume associated with each factory weighted by its value added:

$$Traffic\ Volume\ Index_{i,t} = \sum_{j \in S'_i} \frac{VA_j}{TotalVA'_i} \cdot Traffic_{j,t} \tag{A1}$$

[28] This data is compiled by the National Police Agency using information on traffic volume on general roads collected by prefectural police nationwide and provided by the Japan Road Traffic Information Center. Using the API service provided by the AIGID, traffic volume data can be extracted for any time and location along with the location information of the observation points.

[29] At the same time, it may be capturing information about shipments.

where $Traffic_{j,t}$ is the normalized traffic volume at factory j in month t, setting the average during the observation period (July 2018 to September 2021) to 100, S'_i is the set of factories in sector i for which traffic volume is observed, VA_j is the value added of products produced at factory j, and $TotalVA'_i$ is the total of value added of products produced at factories that belong to S'_i. We can observe traffic volume near about 26% of all factories (in terms of value added).

The Traffic Volume Index and the IIP

The calculated traffic volume index shows a high correlation coefficient of 0.73 with the IIP for the period from July 2019 through September 2021 (Figure 11(a)). The correlation coefficient with the IIP at the industry level is also high in the transport equipment industry and electrical machinery and information and communication electronics equipment industry. The correlation is also relatively high at 0.54 for electronic parts and devices industry, for which the correlation of the mobility index with the IIP is low correlation (0.35) (Figure 11(b), Table 7).

Next, we use a linear regression model to examine whether the traffic volume index could actually improve the nowcasting accuracy of the IIP.[30] Figure 12(a) shows the estimation results when only the traffic volume index is used as an explanatory variable (Specification 1) and when the other alternative data (the mobility index and electricity index) and the IIP Forecast are used in addition to the traffic volume index (Specification 2). We find that the coefficient on the traffic volume index is statistically significant for many industries. In addition, Figure 12(b) compares the adjusted R-squared for Specification 2, which includes the traffic volume index as well as the other alternative data and the IIP Forecast as explanatory variables, and Specification 3, which does not include the traffic volume index but only uses the other alternative data and the IIP Forecast (corresponding to Specification 6 in Table 5). The comparison shows that for all industries, the use of the traffic volume index improves the adjusted R-squared.

The above analysis suggests that the use of alternative data on traffic volume may further improve the accuracy of the nowcasting of the IIP. As mentioned above, while this data is not available in real-time at present, if this or other alternative data of a high public nature become available in real-time in the future, it could potentially be used to improve economic prediction.

[30] Because we cannot assess the accuracy of out-of-sample nowcasts due to the short period for which traffic volume data is available, we check the impact of the use of the traffic volume index on prediction accuracy in-sample.

Table 1. Correlation between the IIP and the Mobility Index at the Industry Level

Industry	Correlation
Mining and manufacturing	0.81
Transport equipment	0.76
Fabricated metals	0.72
Iron, steel and non-ferrous metals	0.67
Electrical machinery, and IC electronics equipment	0.61
Chemicals	0.60
General-purpose and business oriented machinery	0.51
Production machinery	0.46
Pulp, paper and paper products	0.39
Electronic parts and devices	0.35

Note: The table shows the correlation coefficients between the year-on-year rate of change in the IIP and the mobility index (monthly average) at the industry level between Jan. 2018 and Sep. 2021.

Source: Authors' calculations based on data from the Ministry of Economy, Trade and Industry (METI) and Agoop.

Table 2. Correlation between the IIP and the Electricity Index at the Industry Level

Industry	Correlation
Iron, steel and non-ferrous metals	0.76
Mining and manufacturing	0.73
Transport equipment	0.72
Fabricated metals	0.71
Electrical machinery, and IC electronics equipment	0.68
Pulp, paper and paper products	0.68
Production machinery	0.57
General-purpose and business oriented machinery	0.57
Electronic parts and devices	0.53
Chemicals	0.49

Note: The table shows the correlation coefficients between the year-on-year rate of change rate in the IIP and the electricity index (monthly average) at the industry level between Jan. 2018 and Sep. 2021.

Source: Authors' calculations based on data from METI, Organization for Cross-regional Coordination of Transmission Operators (OCCTO), and Japan Meteorological Agency (JMA)

Table 3. Correlation between the IIP and the IIP Forecast at the Industry Level

(a) Following Month IIP Forecast

Industry	Correlation
Iron, steel and non-ferrous metals	0.93
Production machinery	0.92
Transport equipment	0.91
Mining and manufacturing	0.91
General-purpose and business oriented machinery	0.87
Chemicals	0.86
Fabricated metals	0.85
Electrical machinery, and IC electronics equipment	0.84
Pulp, paper and paper products	0.71
Electronic parts and devices	0.68

(b) Current Month IIP Forecast

Industry	Correlation
Transport equipment	0.97
Mining and manufacturing	0.97
Iron, steel and non-ferrous metals	0.96
Production machinery	0.95
Fabricated metals	0.94
Electrical machinery, and IC electronics equipment	0.91
Pulp, paper and paper products	0.91
Chemicals	0.88
General-purpose and business oriented machinery	0.87
Electronic parts and devices	0.80

Note 1: The tables show the correlation coefficients between the year-on-year rate of change in the IIP and the IIP Forecast at the industry level between Jan. 2018 and Sep. 2021.
Note 2: The bias-adjusted IIP Forecast is used.
Source: Authors' calculations based on data from METI.

Table 4. Chronological Relationship of Data Used and Nowcasting Models

		IIP Forecast/IIP Schedule	Nowcasting models in this paper
Previous month $(t-1)$	Until the 10th of the month	Survey responses for the IIP Forecast for the following month (for month t) are submitted (deadline on the 10th).	
	End of the month	The IIP Forecast for the following month (month t) is published.	
Current month (t)	7th of the month		Nowcasting models for the 1st week
	Until the 10th of the month	Survey responses for the IIP Forecast for the current month (for month t) are submitted (deadline on the 10th).	
	14th of the month		Nowcasting models for the 2nd week
	21st of the month		Nowcasting models for the 3rd week
	End of the month	The IIP Forecast for the current month, (month t) is published.	
			Nowcasting models for the end of the month
Following month $(t+1)$	End of the month	The IIP for month t is published.	

Table 5. Specification List

	Specification (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mobility Index	✔		✔	✔		✔	✔		✔
Electricity Index		✔	✔		✔	✔		✔	✔
IIP Forecast				✔	✔	✔	✔	✔	✔
Mobility Index (lagged by one month)							✔		✔
Electricity Index (lagged by one month)								✔	✔
IIP (lagged by one month)							✔	✔	✔

Note 1: We use the following month IIP Forecast in the 1st to 3rd weeks and the current month IIP Forecast at the end of a month.
Note 2: Since the IIP is not published until the end of the following month, we use the current month IIP Forecast for "IIP (lagged by one month)" in the 1st to 3rd weeks.

Table 6. Best Model for Each Industry and Point in Time

Industry	1st Week	2nd Week	3rd Week	End of Month
Mining and manufacturing	RF Mix (7)	GB Mix (6)	Lin Reg (4)	GB Mix (9)
Iron, steel and non-ferrous metals	GB Mix (7)	GB Mix (7)	GB Mix (9)	GB Mix (6)
Fabricated metals	GB Mix (7)	GB Mix (6)	GB Reg (5)	GB Mix (6)
Production machinery	Lin Reg (6)	RF Mix (5)	RF Mix (5)	GB Reg (8)
General-purpose and business oriented machinery	GB Mix (4)	Lin Reg (4)	Lin Reg (4)	Lin Reg (4)
Electronic parts and devices	RF Reg (8)	RF Reg (8)	RF Reg (8)	Lin Reg (9)
Electrical machinery, and IC electronics equipment	Lin Reg (5)	Lin Reg (5)	RF Mix (6)	GB Mix (9)
Transport equipment	GB Mix (6)	RF Mix (6)	RF Mix (4)	RF Mix (4)
Chemicals	GB Reg (8)	GB Reg (8)	GB Reg (9)	GB Mix (7)
Pulp, paper and paper products	RF Reg (8)	GB Reg (5)	GB Reg (5)	GB Mix (6)

Note 1: The table shows the nowcasting models with the smallest RMSE for a particular industry and point in time.
Note 2: "Lin Reg" refers to the linear regression model, "RF Reg" and "GB Reg" refer to the random forest regression model and the gradient boosting regression model, and "RF Mix" and "GB Mix" refer to the mixed model using random forests and that using gradient boosting, respectively. The numbers in parentheses denote the combination of variables used (see Table 5).
Note 3: Shaded cells denote that the RMSE of the model is greater than that of the IIP Forecast.

Table 7. Correlation between the IIP and the Traffic Volume Index at the Industry Level

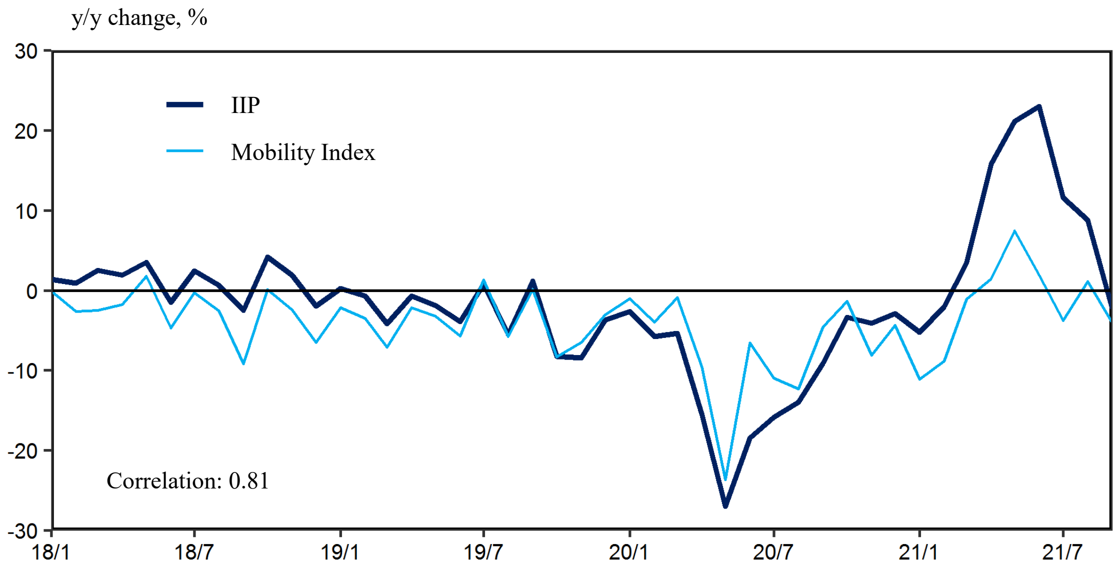
Industry	Correlation
Transport equipment	0.81
Mining and manufacturing	0.73
Electrical machinery, and IC electronics equipment	0.69
General-purpose and business oriented machinery	0.68
Electronic parts and devices	0.54
Fabricated metals	0.53
Production machinery	0.44
Pulp, paper and paper products	0.43
Iron, steel and non-ferrous metals	0.39
Chemicals	0.17

Note: The table shows the correlation coefficients between the year-on-year rate of change in the IIP and the traffic volume index (monthly average) at the industry level between Jul. 2019 and Sep. 2021.

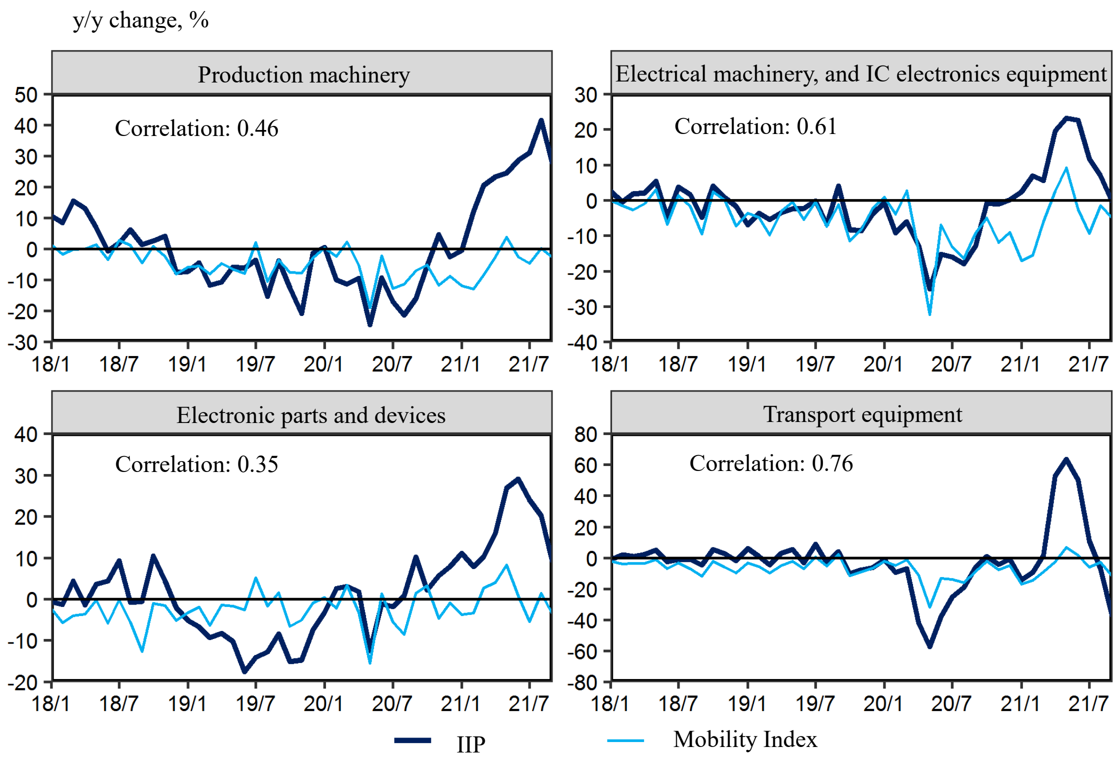
Source: Authors' calculations based on data from METI and the Association for Promotion of Infrastructure Geospatial Information Distribution (AIGID).

Figure 1. The IIP and the Mobility Index

(a) Manufacturing and mining

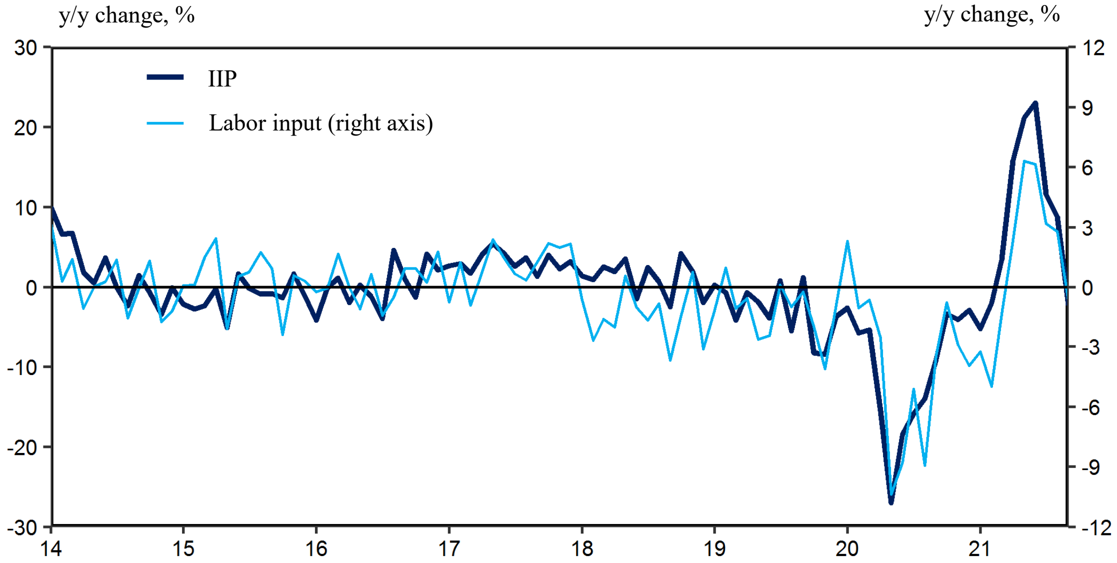


(b) Major Industries



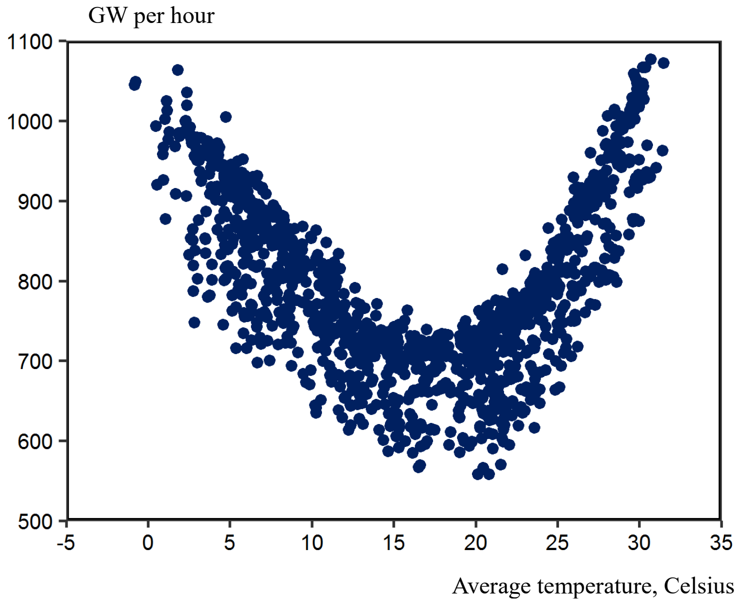
Note: The mobility index is its monthly average values.
Source: Authors' calculations based on data from METI and Agoop.

Figure 2. The IIP and Labor Input



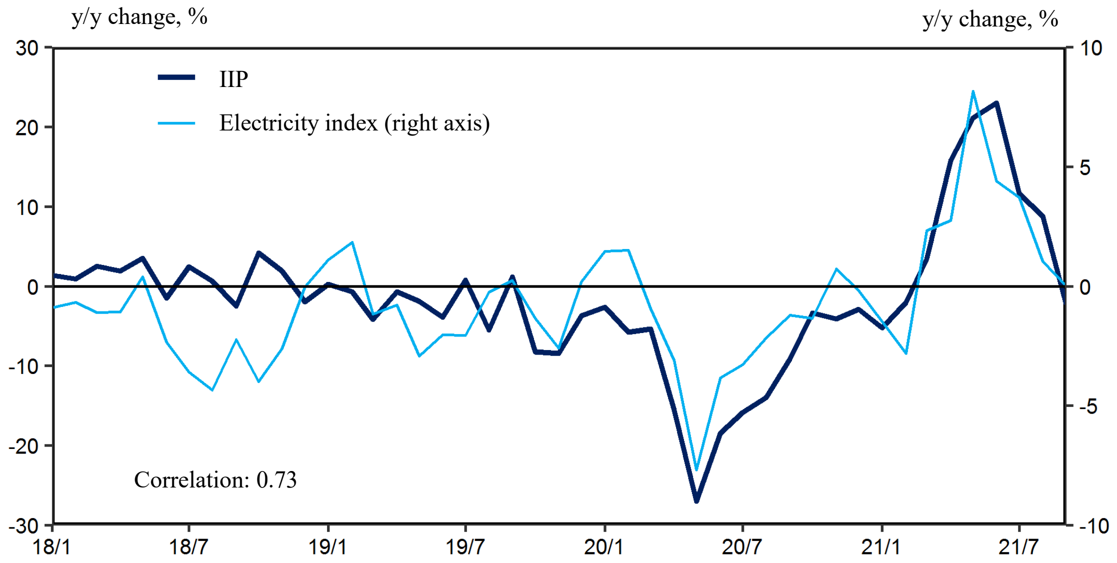
Note 1: Labor input equals the total hours worked per worker times the number of employed persons.
Note 2: The IIP is that for manufacturing and mining, while labor input is that in the manufacturing sector.
Source: Authors' calculations based on data from METI and Ministry of Health, Labour and Welfare.

Figure 3. Temperature and Electricity Demand



Note: The figure shows the daily electricity demand and average temperature in Tokyo between Jan. 2018 and Sep. 2021.
Sources: OCCTO, JMA.

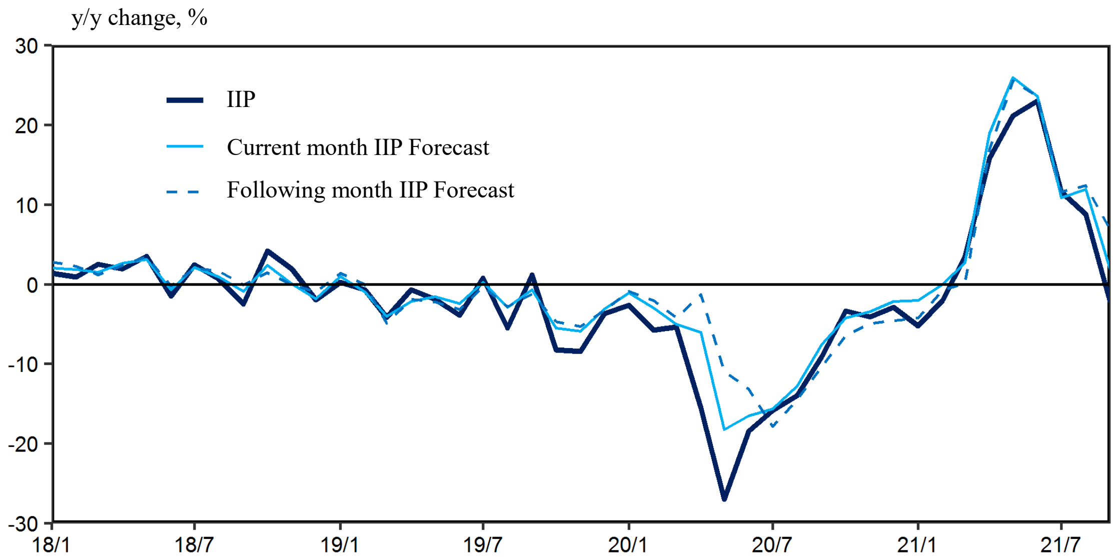
Figure 4. The IIP and the Electricity Index



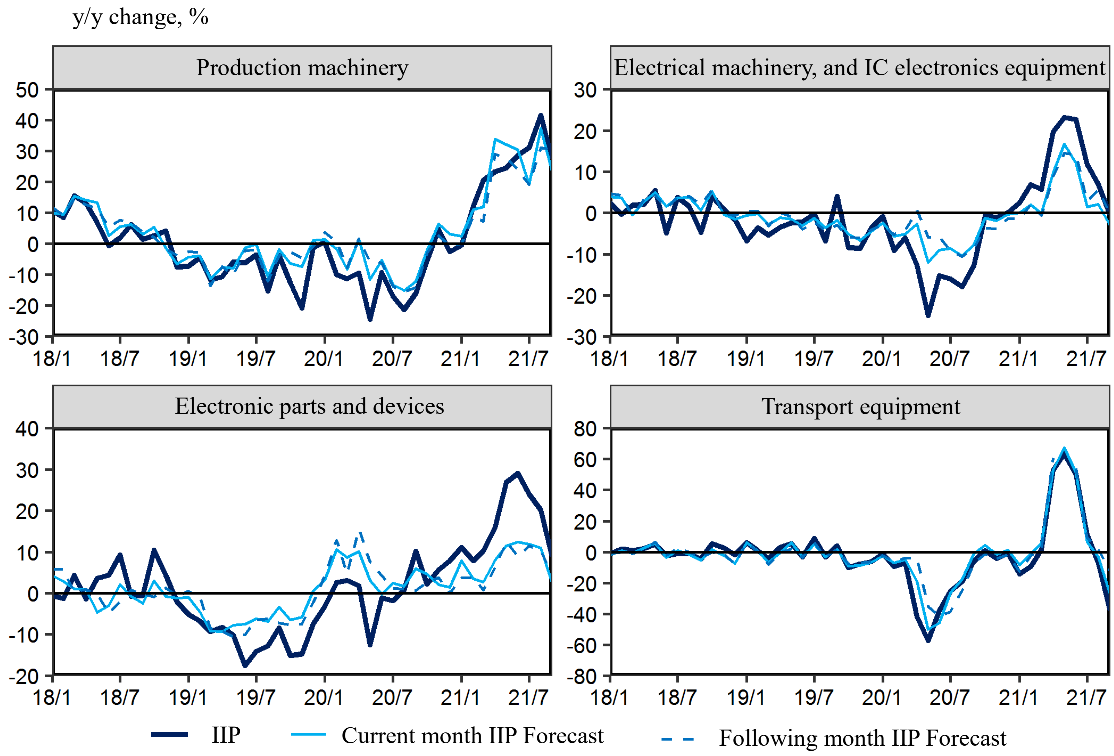
Note: The electricity index is its monthly average values.
Source: Authors' calculations based on data from METI, OCCTO, and JMA.

Figure 5. The IIP and the IIP Forecast

(a) Manufacturing and mining

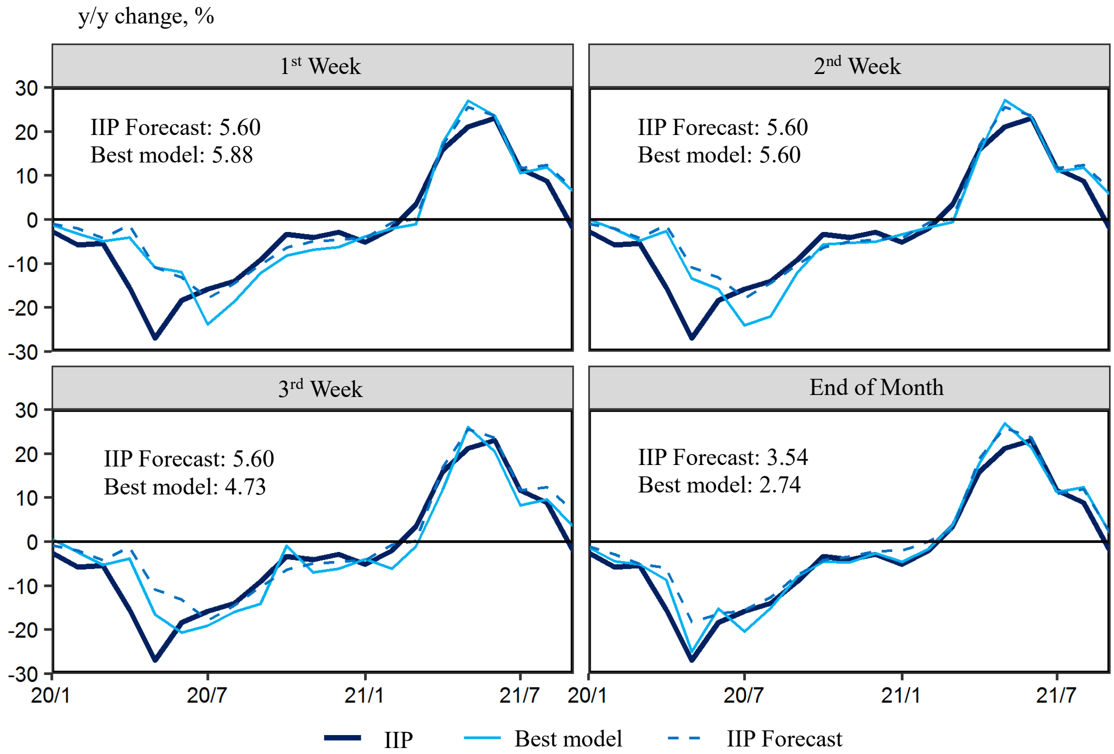


(b) Major industries



Note: The bias-adjusted IIP Forecasts are used.
Source: Authors' calculations based on data from METI.

Figure 6. Nowcasting Results (Manufacturing and Mining)

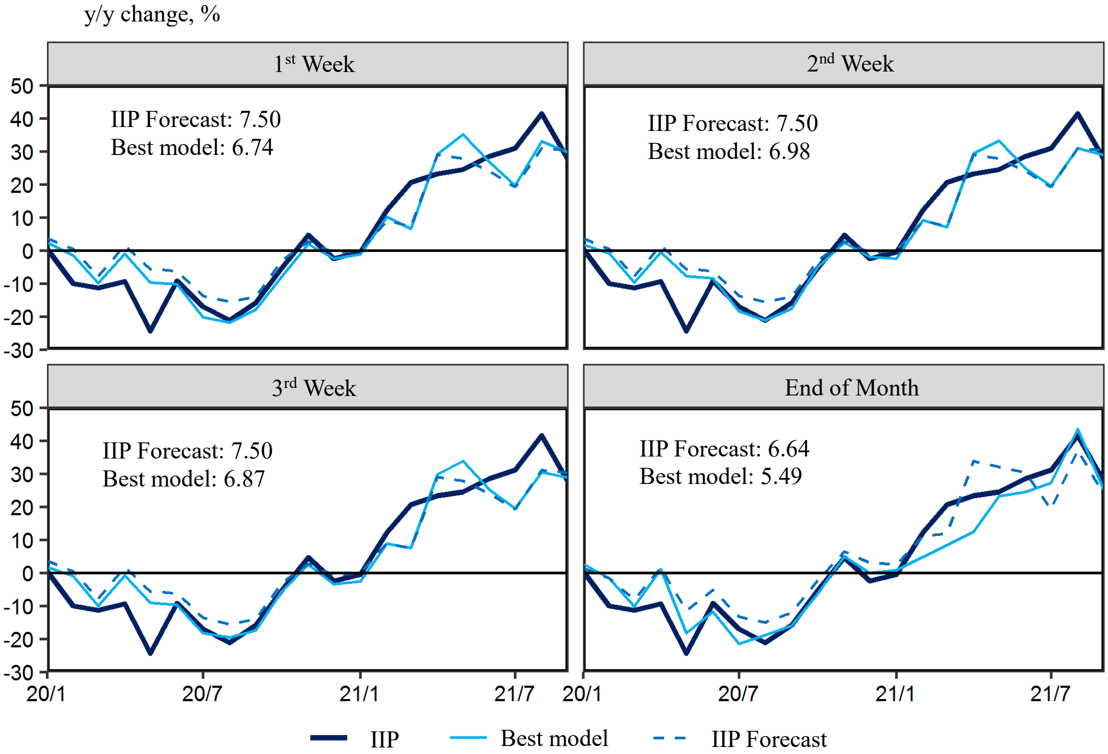


Note 1: "IIP Forecast" shows the current month IIP Forecast for the 1st to the 3rd week and the following month IIP Forecast at the end of the month.

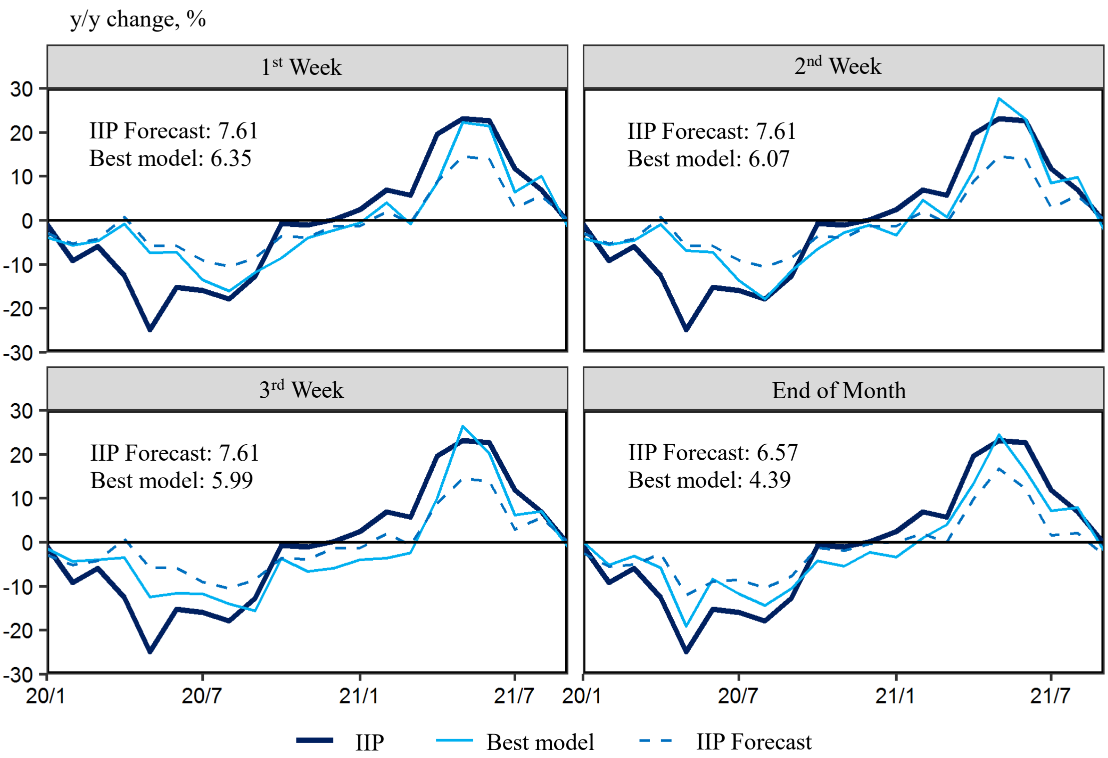
Note 2: The numbers in the figure are RMSEs.

Figure 7. Nowcasting Results (Major Industries)

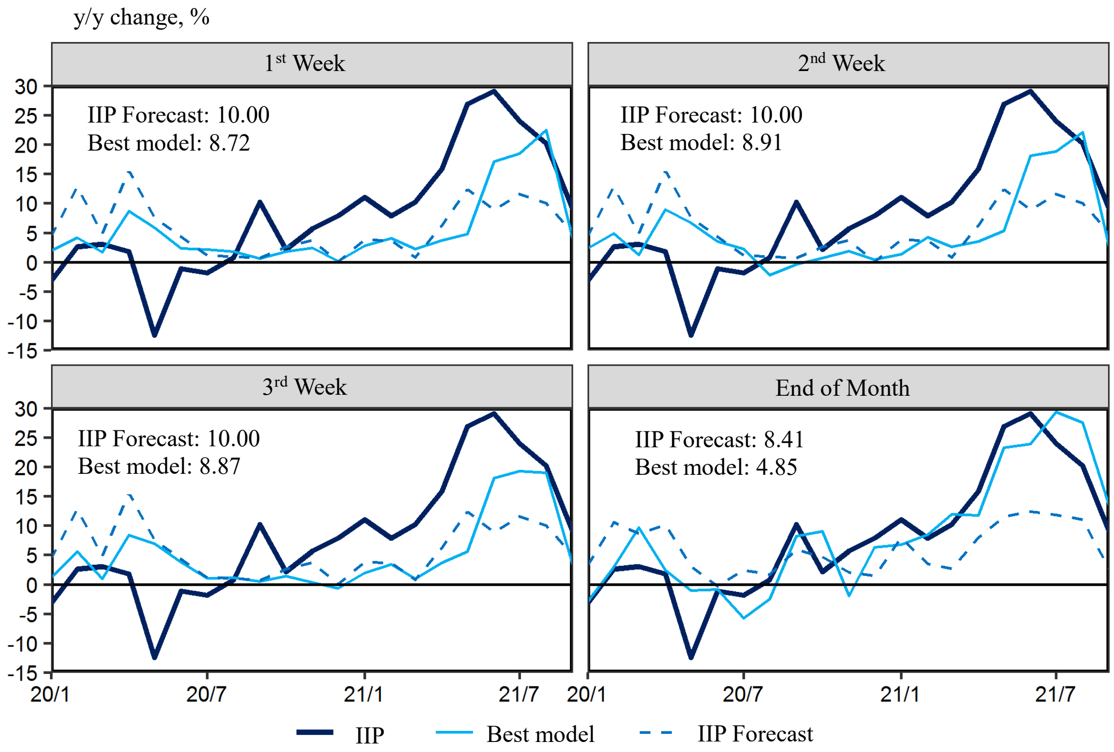
(a) Production machinery



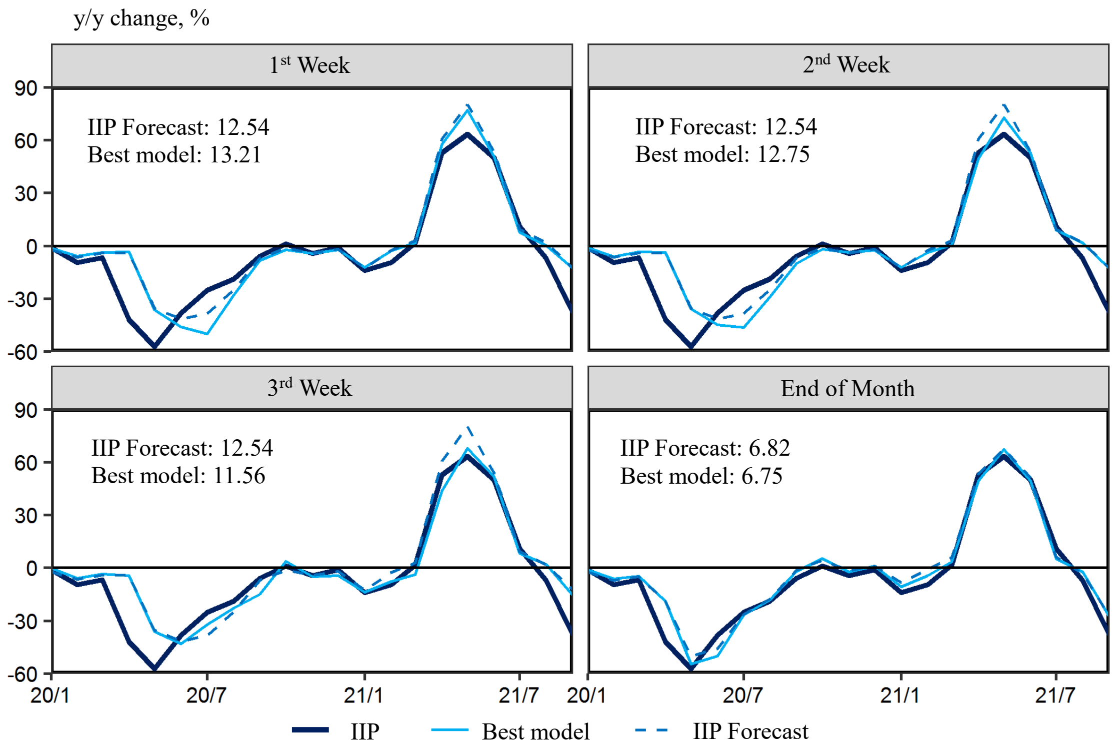
(b) Electrical machinery, and IC electronics equipment



(c) Electronic parts and devices

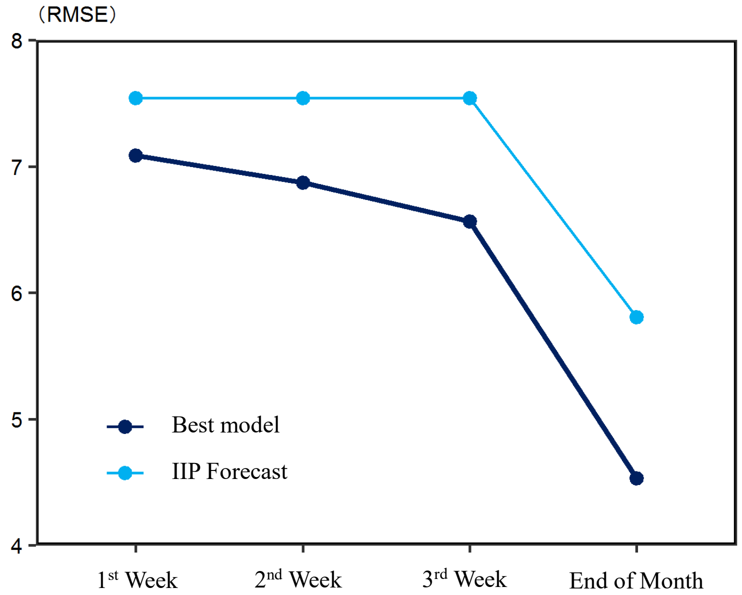


(d) Transport equipment



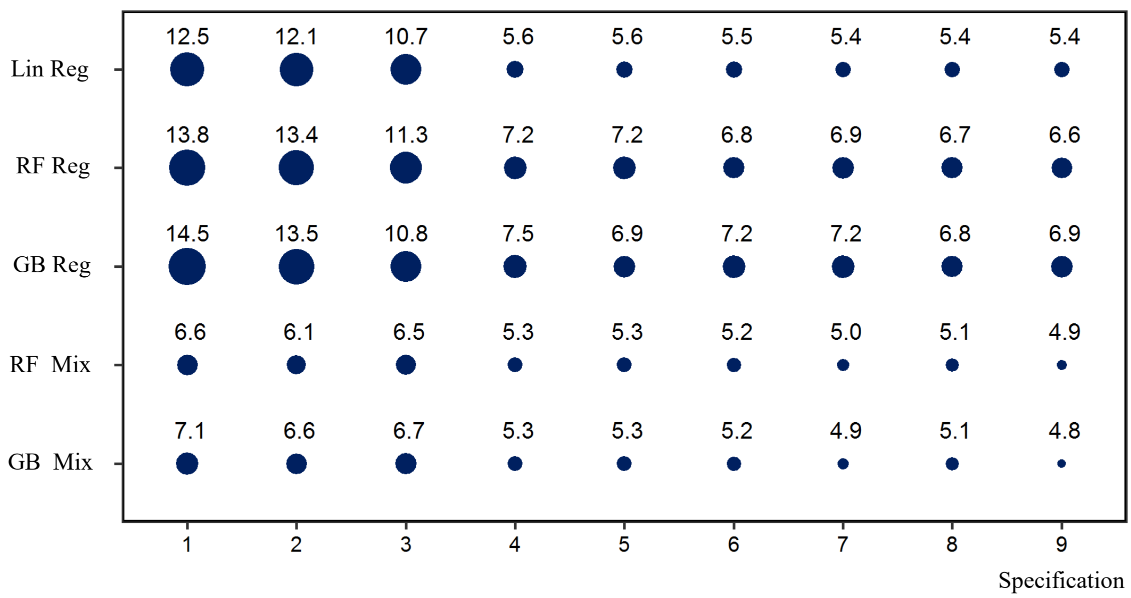
Note 1: "IIP Forecast" shows the current month IIP Forecast for the 1st to the 3rd week and the following month IIP Forecast at the end of the month.
Note 2: The numbers in the figure are RMSEs.

Figure 8. RMSE of Nowcasting Models by Time



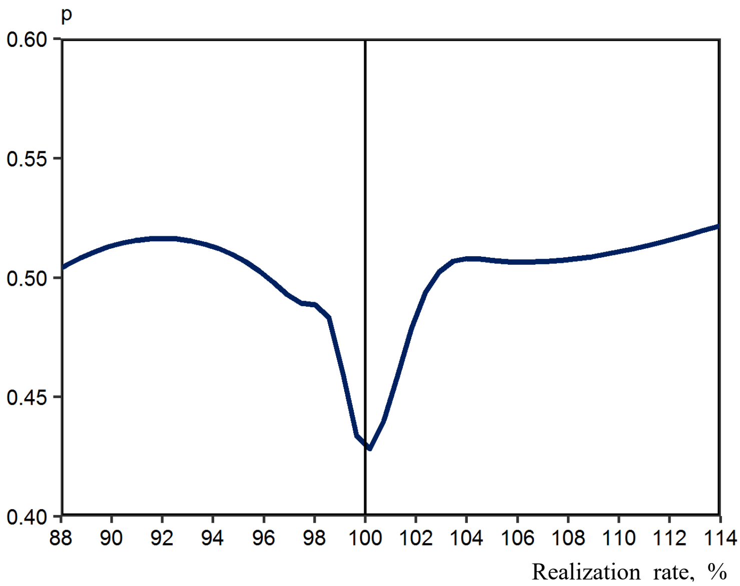
Note 1: "IIP Forecast" shows the current month IIP Forecast for the 1st to the 3rd week and the following month IIP Forecast at the end of the month.
Note 2: The figure shows the cross-industry average RMSE.

Figure 9. RMSE of Nowcasting Models by Model Type



Note 1: "Lin Reg" refers to the linear regression model, "RF Reg" and "GB Reg" refer to the random forest regression model and gradient boosting regression model, and "RF Mix" and "GB Mix" refer to the mixed model using random forests and that using gradient boosting, respectively. The numbers on the x-axis correspond to the specification numbers shown in Table 5.
Note 2: The numbers in the figure are the cross-industry average RMSE of each model and specification at the end of the month (the size of the bubbles corresponds to the RMSE)

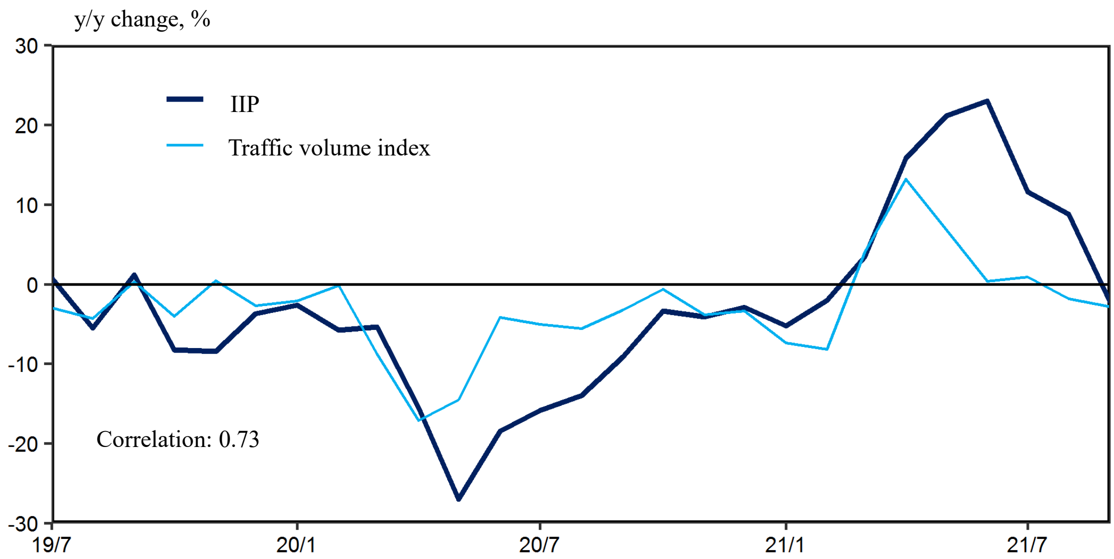
Figure 10. "Realization Rate" of the IIP Forecast and p



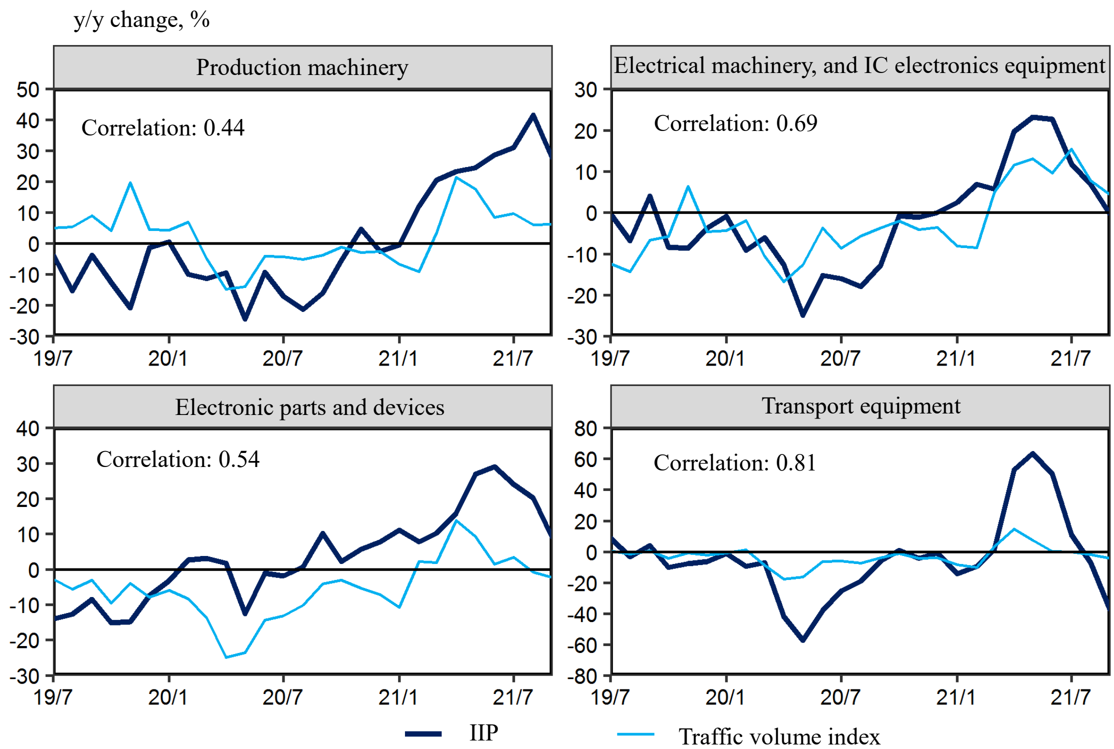
Note 1: The figure is based on the estimation results of the two models (the random forests and gradient boosting models) and the nine specifications for all industries.
Note 2: The line represents the fitted curve using a general additive model.

Figure 11. The IIP and the Traffic Volume Index

(a) Manufacturing and mining



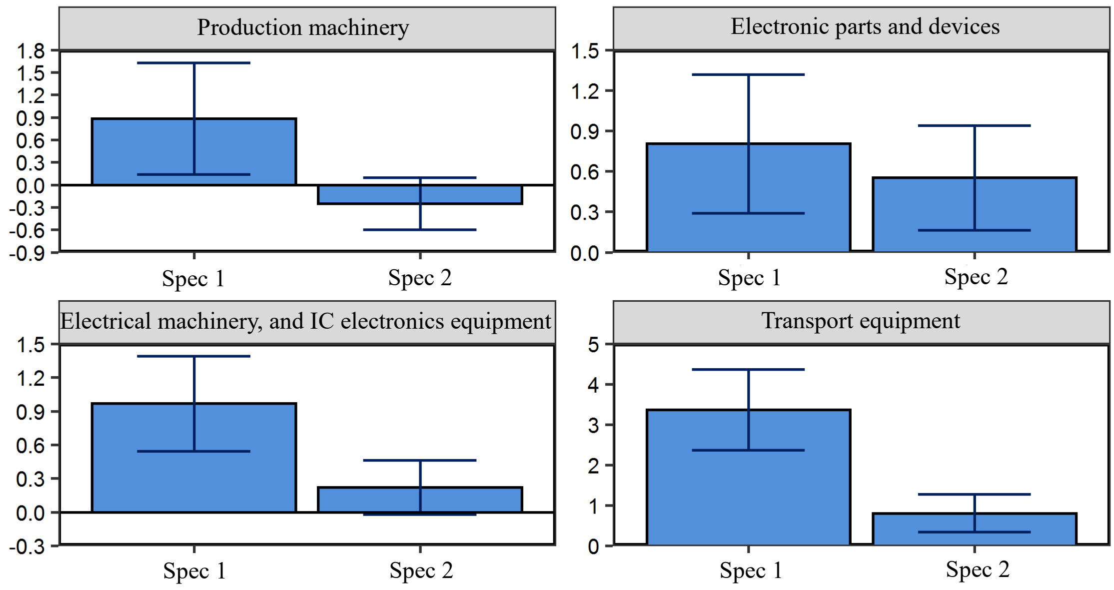
(b) Major industries



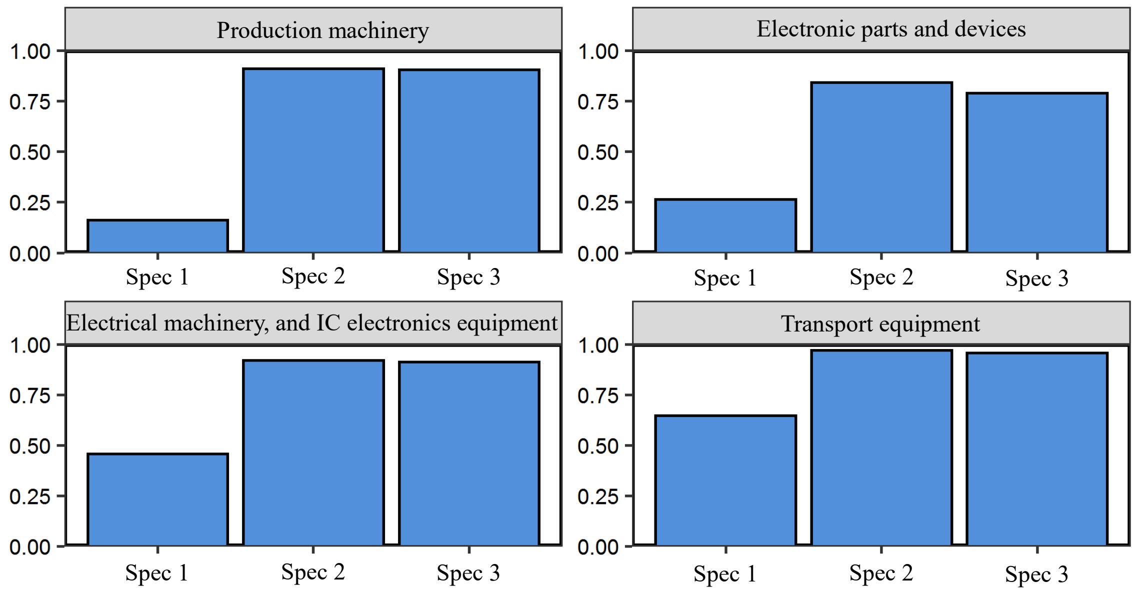
Source: Authors' calculations based on data from METI and AIGID.

Figure 12. The IIP and the Traffic Volume Index (OLS Estimation Results)

(a) Estimated coefficients on the traffic volume index and their confidence intervals



(b) Adjusted R-squared



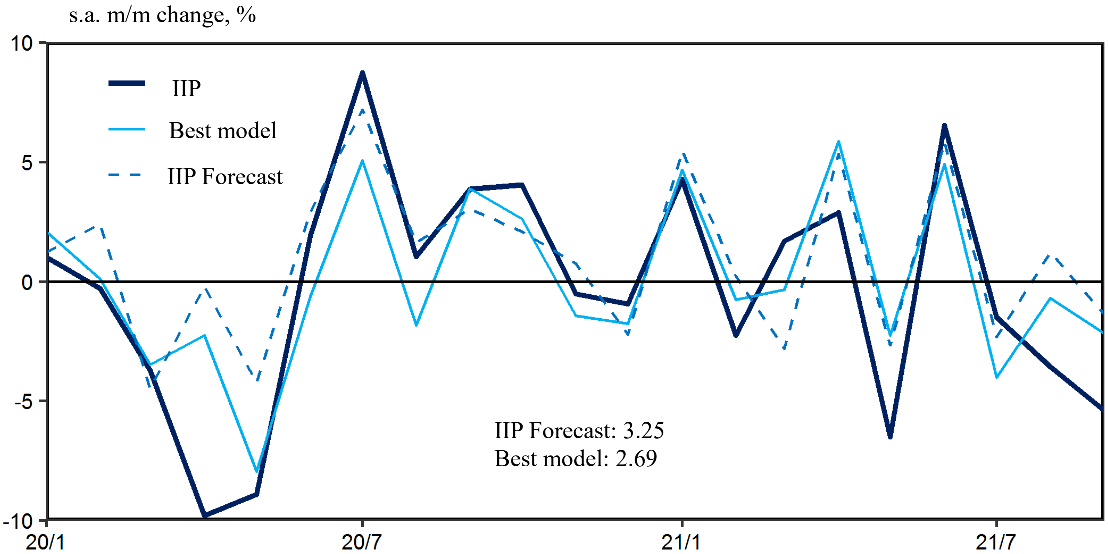
Note 1: "Spec 1" is a regression model in which the explanatory variable is the traffic volume index. "Spec 2" is a regression model in which the explanatory variables are the mobility index, the electricity index, the IIP Forecast, and the traffic volume index. "Spec 3" is a regression model in which the explanatory variables are the mobility index, the electricity index, and the IIP Forecast. The dependent variable in all the specifications is the IIP.

Note 2: The error bands represent the 95% confidence intervals.

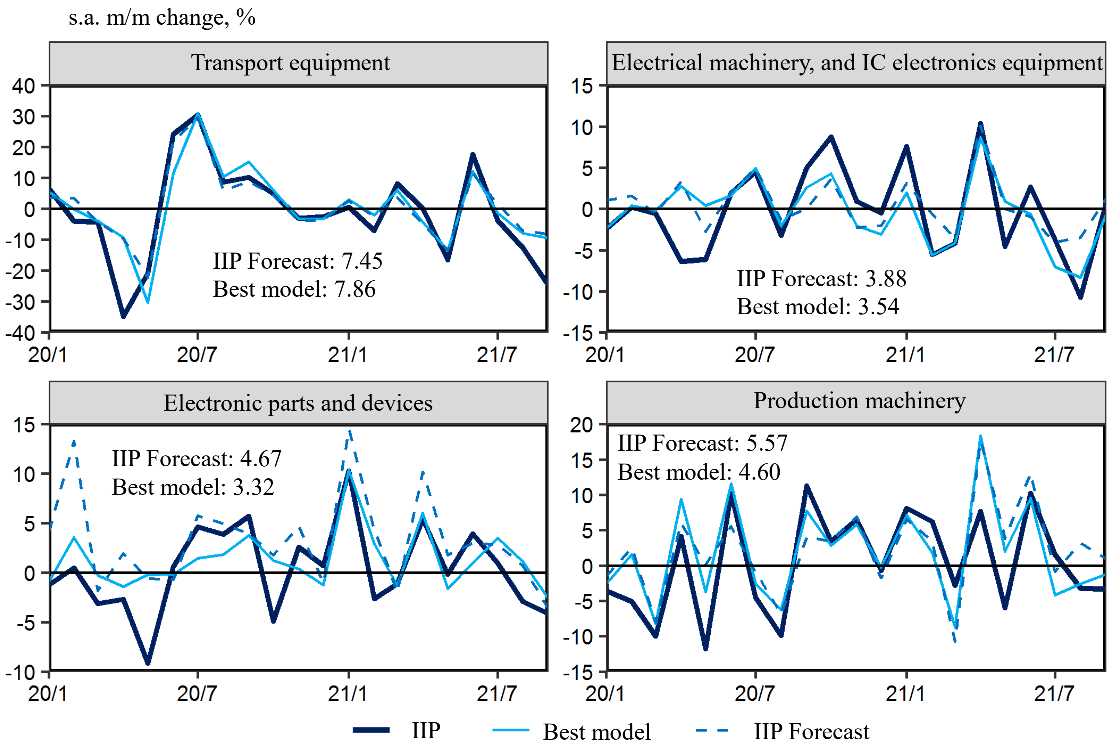
Appendix Figure 1. Estimation Results of the Nowcasting Models

(Seasonally Adjusted)

(a) Manufacturing and mining



(b) Major industries



Note 1: The numbers in the figures are the RMSEs.
Note 2: "IIP Forecast" shows the current month IIP Forecast.