J. Korean Soc. Ind. Appl. Math. Vol.26, No.4, 280–295, 2022 http://doi.org/10.12941/jksiam.2022.26.280

IMPROVING GLOBAL SUPPLY CHAIN RISK IDENTIFICATION USING RCF

MYUNGHYUN JUNG[1,†], SEYEON LEE[1], MINJUNG GIM[1], HYUNGJO KIM[2], AND JAEHO LEE[2]

[1]DEPARTMENT OF INNOVATION CENTER FOR INDUSTRIAL MATHEMATICS, NATIONAL INSTITUTE FOR MATHEMATICAL SCIENCES, SOUTH KOREA

Email address: †mhjung07@nims.re.kr

[2]GVC CENTER, INDUSTRIAL POLICY DIVISION, KOREA ASSOCIATION OF MACHINERY INDUSTRY, SOUTH KOREA

ABSTRACT. This paper contains an introduction to industrial problems, solutions, and results conducted with the Korea Association of Machinery Industry. The client company commissioned the problem of upgrading the method of identifying global supply risky items. Accordingly, the factors affecting the supply and demand of imported items in the global supply chain were identified and the method of selecting risky items was studied and delivered. Through research and discussions with the client companies, it is confirmed that the most suitable factors for identifying global supply risky items are 'import size', 'import dependence', and 'trend abnormality'. The meaning of each indicator is introduced, and risky items are selected using export/import data until October 2022. Through this paper, it is expected that countries and companies will be able to identify global supply risky items in advance and prepare for risks in the new normal situation: the economic situation caused by infectious diseases such as the COVID-19 pandemic; and the export/import regulation due to geopolitical problems. The client company will include in his report, the method presented in this paper and the risky items selected by the method.

1. INTRODUCTION

Uncertainty in the global supply chain is growing due to the recent events that have a significant impact on the global economy: the 2019 Korea-Japan trade dispute; the 2020 COVID-19 pandemic; the 2021 urea crisis; and the 2022 Russia-Ukraine war. In this uncertain situation, major countries find and monitor risky items through various methods to identify and prepare for the vulnerability in supply chains in advance. This paper contains the problem solving conducted with the GVC (Global Value Chain) Center of the Korea Association of Machinery Industry, a client company, and deals with how to select global supply chain risky items.

The industrial problem requested by the client company is to verify the method of identifying risky items used in the GVC center and to present new methods improved in a mathematical

Received by the editors December 2 2022; Revised December 14 2022; Accepted in revised form December 15 2022; Published online December 25 2022.

2000 *Mathematics Subject Classification.* 93B05.

Key words and phrases. Global supply chain risky item identification, Import size, Import dependence, Entropy, Time series trend anomaly detection, RCF.

† Corresponding author.

approach. First, I will introduce the recent global supply chain situation by referring to Deloitte Insights No.22 in 2022 [1].

As the international situation has changed rapidly in the past two to three years, global supply chain threat issues have arisen that do not properly import raw materials or intermediate goods needed for production. The representative issues are as follows.

- COVID-19 Pandemic: The last COVID-19 pandemic has caused economic lockdowns in each country. As a result, there were problems in the supply and demand of raw materials and intermediate goods. The paralysis of logistics system and labor shortage occurred.
- Geopolitical problems: The U.S.-China conflict, and Russia's war of aggression in Ukraine disrupted supply and demand for raw materials, food, parts and finished goods, creating chaos in the global supply chain.

Due to frequent confusion in the supply chain, companies are responding to this extreme volatility by defining the economic situation as 'new normal'. Companies are investing in a variety of research and infrastructure to efficiently manage supply chains and operate reliably in new normal situations. For example, Korea has begun to diversify its import suppliers for items that have become impossible to import due to the blockade caused by COVID-19 or trade disputes between Korea and Japan. In addition, the companies in the machinery industry use techniques of AI and data analysis to introduce an intelligent factory system that analyzes and interprets manufacturing data, and make efforts to increase efficiency through production optimization.

The GVC center of client company is engaged in various activities to prepare for the risks of such an import supply chain in advance. Typically, the center publishes a "GVC Supply Chain Risky Item Report" [2] every month to identify items that need monitoring.

2. Related Research

Methods for selecting global supply chain risky items have recently been studied in various ways. Min and Lee [3] analyzed the vulnerability to the Korean economy by dividing it into domestic and foreign. The vulnerability of global supply chain is analyzed using the IMF's network analysis theory, and that of domestic supply chain is analyzed using the regional supply chain vulnerability check methodology proposed by the EU Commission. As a result, the vulnerability of global supply chain was found to be due to the high dependence on key trading partners such as China, the United States, and Germany and the clustering phenomenon between trading countries. In particular, the dependence on China has increased over the past 10 years since 2012, as China's influence in the global trade has expanded. In the case of domestic supply chains, raw materials were very vulnerable, but capital goods showed relatively low vulnerability. This analysis result is judged to be significant because it uses the import dependence using trade status and export/import amount. However, in this study, the trend of import dependence and export/import amount over long period units (minimum 1 year to maximum 4 years) was analyzed to determine vulnerability. Therefore, it was difficult to apply the above analysis method to select risky items for the import supply chain every month.

In order to analyze the vulnerability of supply chain in the Korean industry, Kim et al. [4] divided items into consumer goods, intermediate goods, and capital goods, and compared and analyzed the vulnerability to China in Korea, the United States, and Japan. Among Korean imports to China, intermediate goods were found to be mainly weak. This turned out to be due to the division of labor between Korea and China. This analysis method is useful for identifying vulnerabilities to target countries by item, so it can prepare for vulnerable items in advance. However, by simply calculating the import dependence for each item in 2020, if it is more than 50%, it was selected as an item of interest, but it did not take into account changes in import dependence, import volume, amount, or trends in import unit prices. In addition, it contained only analysis of the United States, Japan, and China, so it was limited to apply it directly to the global supply chain.

3. Proposed Method

This section introduces export/import data provided by the client company to solve the problem. In addition, we introduce the conventional method of identifying global supply chain risky items at the GVC center first, and introduce new item identification methods that have been discussed and advanced by research institutes and companies later. The trade-related knowledge and the definition and judgment of risky items were provided by the company. The institute interpreted them through data search and developed modeling and machine learning methods to reflect the criteria for risky items. We decided which factors are more important in determining global supply chain risky items, and presented a mathematical method for detecting time series trends using the characteristics of time series trade data.

3.1. **Data description.** The client company provided export/import data of items imported in Korea to solve the problem. The contents of data are as follows.

- Export/Import amount (unit: $) and weight (unit: kg) by item
- List of countries which exported to Korea by item and weight of items imported from each country
- Period: January 2017 to October 2022

Table 1 shows an import data sample. The item name is 'Lithium polymer battery', and 'v' and 'w' in the column name mean the total amount and weight, respectively, of the item imported in the target year and month. The 'iso_n' column means the country from which Korea imports. That is, in the first row of the import data, the import amount (v) and the import weight (w) of the lithium polymer battery imported from China every month are recorded.

TABLE 1. Import data for lithium polymer battery.

index	iso_n	v2201	v2202	...	w2201	w2202	...
0	China	84,999,667	70,504,831	...	2,856,201	2,462,474	...
1	Vietnam	2,605,022	5,208,289	...	30,349	65,561	...
2	Japan	316,104	2,736,576	...	3,253	143,987	...

The visualization of time series data of the import amount, weight, Trade Specialization Index (TSI), entropy, and the import ratio of the largest importing country is as the following Fig. 1. TSI, entropy, and the import ratio of the largest importing countries are made by processing import data with export one, and detailed explanations will be given below.

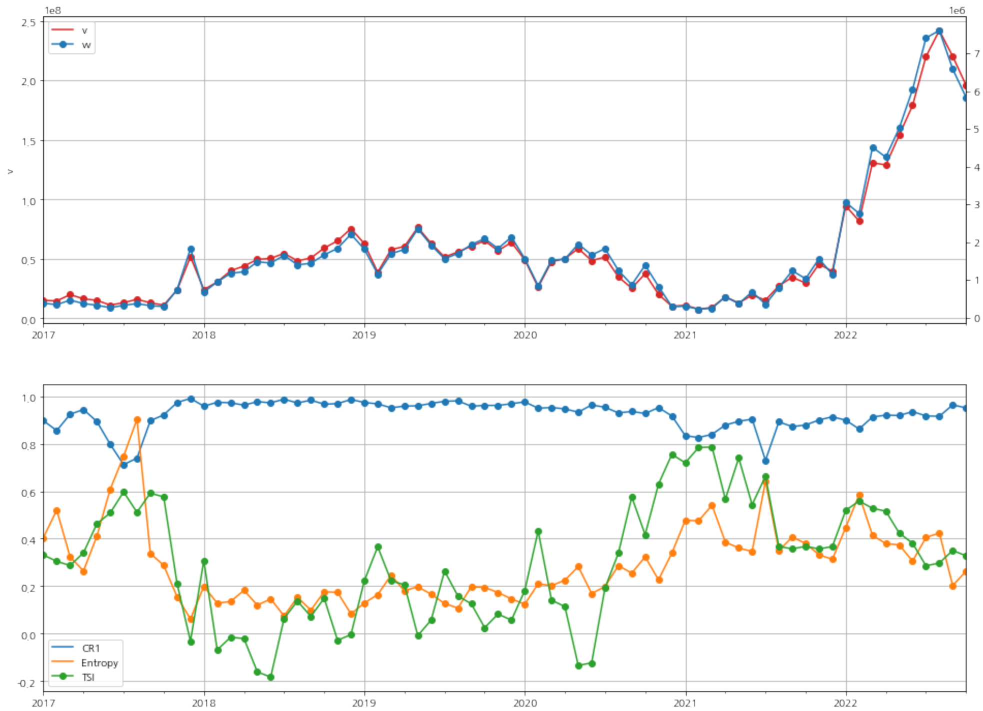


FIGURE 1. Lithium polymer battery.

3.2. **Conventional methodology for risky item identification.** We will introduce the definition of global supply chain risky items and the risky item identification method currently in use by the client company. The GVC center focuses on imports rather than exports. When items are difficult to find alternative import countries when imports from certain countries are suspended due to problems in the global supply chain or the import amount and weight of items differ significantly from past patterns, then the items are of interest. The GVC center thinks that the surge in domestic demand could significantly change the amount of imports or unit prices of imports, and cause problems in supply and demand. Therefore, the GVC Center determines items that need monitoring with the size of import and import growth by item, import surge, import dependence, and TSI in the export/import data. If all specific criteria for each indicator are met, the item is designated as a global supply chain risky item. Let's look at each element.

- The size of import and import growth: When the sizes of import and import growth are more than 1 million dollars. This is an indicator that is examined to prevent excessive risk alerts and to focus on items with a certain size or larger of imports.
- Increase in import: When the increase in import in the target month compared to the same month last year is within the top 1% of distribution of monthly variances in the past 55 months.
- Import dependence: This means the percentage of import from a specific country in the target month. The 50-75% dependence on import from a specific country implies a cautionary item, and the 75% above dependence a vulnerable item.
- TSI: TSI is defined as ((Export amount)-(Import amount))/((Export amount)+(Import amount)). When TSI is negative, the item is selected as a cautionary item. When TSI is less than -0.5, the item is selected as a vulnerable item.

Items satisfying at least one criterion is of interest. Items that meet all four criteria are selected as global supply chain risky items. Let's look at the Fig. 2.

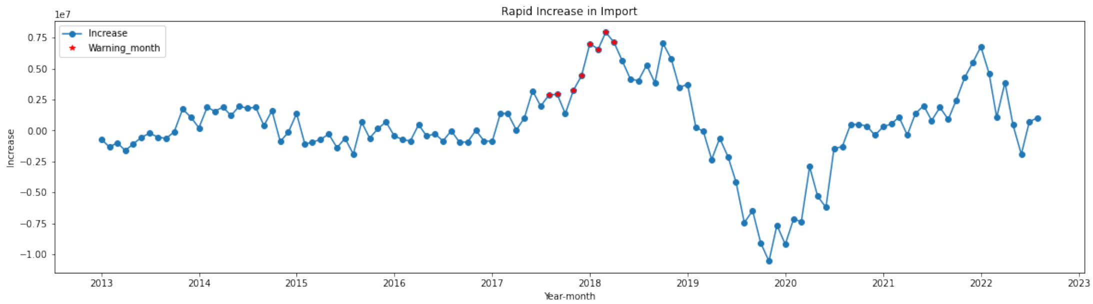


FIGURE 2. The variance in import for etching gas.

This is a graph showing the variance in import for etching gas from January 2013 to August 2022 compared to the same month last year. x-axis means time, and y-axis the amount of import from a month minus that from the same month last year. The red dot indicates that the amount on the target month is within the top 2% (a little more flexible than the standard of client company) in the distribution of monthly variances over the past 55 months. The item on the month corresponding to the red dot is selected as a item of interest for the "increase in import" indicator.

The institute judges that the method of client company needs the following improvements, although it is useful for identifying import-related risky items.

1) Reliability of criteria by indicator: The criteria for each indicator in the method involve subjective and empirical judgment which may reduce reliability of identified risky items.
2) Not considering time series trends: In the case of 'import dependence' and 'TSI' indicators, only data from the target month is used to determine whether the item is risky

on the month. Since it is the time series data, it seems necessary to analyze time series trends to detect the trend abnormality.

3) Not considering import dependence trend: In the case of 'Import dependence' indicators, '75% or more' criterion is considered only for the country with the largest import dependence. However, even if an item is currently identified as one of interest, there are cases where the import dependence is decreasing, but this is not considered. In addition, the institute thinks that it is necessary to consider the second or lower proportion of import.

For example, let's take a look at the Fig. 3 of 'increase in import' indicator for the etching gas item, already shown above.

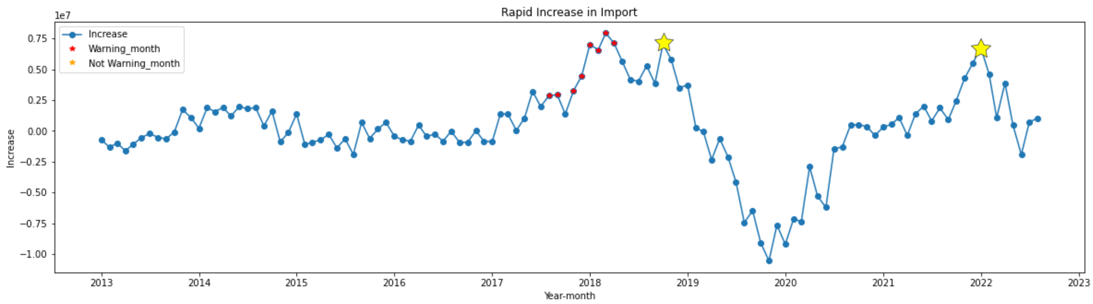


FIGURE 3. The increase in import.

The red dot shows that the month is within the top 2% (a little more flexible selection of the standard) in the distribution of monthly increases and decreases over the past 55 months. The amounts of variance in import in October 2018 and January 2022 expressed in yellow dots are also significant. But the months cannot be risk-identified even though the identification standard is a flexible one of 2%. This is due to the masking problem of not recognizing the yellow dot as a abnormal value due to the high value (red dot) that occurred before October 2018. In other words, it is not clear why the 'within the top 1%' standard is selected as the criterion for identifying risky items in 'the past 55-month distribution'.

3.3. **Enhanced methodology for risky item identification.** This chapter will deal with the problem-solving methods presented by the institute to the client company. Analysis indicators are selected by referring to the client company's advice, papers and reports related to the trade and statistics.

- **Size of import**: Items with the average amount of import for 12 months, including the target month, greater than a certain amount, are selected as items of interest. If the import amount is less than a certain amount, it is excluded because it is not influential among the entire trade items. The "variance in import" is not considered because it could be affected by the value in the same month last year which is not normal.

- **Import dependence**: If the proportion of import to a specific country to total import amount exceeds a certain standard, or if the import dependence tends to be concentrated on a specific country over the past six months, it is selected as item of interest. As a method of measuring the trend of import dependence, the entropy of import ratio by importing country is used. Entropy can use the import ratio of other countries as well as that of largest importing country. The change of entropy along the import ratio is greater than other indicators, so it is adequate to look at the trend of import dependence.
- **Trend anomaly analysis**: The total amount of import, import weight (scale), and TSI values of last four months, including the target month, are exploited with a sliding window method. This is because information and time series characteristics of adjacent months can be used only when trends are analyzed, not just by looking at indicators on specific months.

  If the value on the last four months is higher than the average trend, it is selected as a risky item in the supply chain. The reason for comparing it with the average trend is that if abnormal values exist in the past, this may lead to a masking problem that the abnormal value on the target month is identified as a normal value.

3.4. **Detailed description of enhanced methodology.**

3.4.1. *The import dependence and Entropy.* We measured the import dependence for items using the import ratio of largest importing country and entropy. First of all, like the conventional GVC center method, if the highest import ratio of an item exceeds a certain value (e.g., 90%), it was selected as an item of interest. In addition, if the import ratio of largest importing country is moderately high (e.g., 75% to 90%), then it has been determined whether the import dependence trend over the past six months is focused on specific countries. Here, we measured the import dependence using entropy, which is described later, to investigate the import dependence trend. To measure the variance of entropy trend, we conducted a least square linear regression analysis, which minimizes the sum of squared residual between observed data and approximated output, on the time series data of last 6 months. More precisely, given n input-observation pairs (x_i, y_i), we want to find parameters $(\hat{\alpha}, \hat{\beta})$ which satisfies

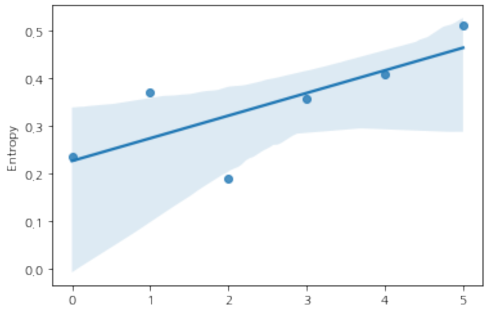
$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha,\beta\in\mathbb{R}}{\arg\min} \sum_{i=1}^{n} (\alpha x_i + \beta - y_i)^2 := \underset{\alpha,\beta\in\mathbb{R}}{\arg\min} E(\alpha, \beta).$$

where, $\hat{\alpha}$ is slope and $\hat{\beta}$ is bias. In order to find parameters, it is enough to solve the following:

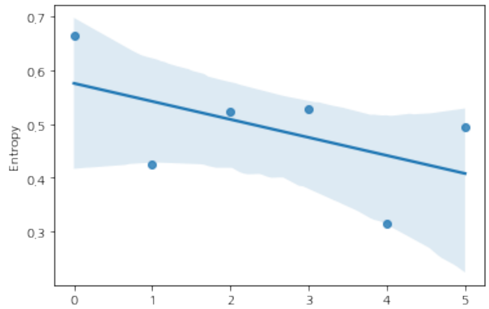
$$\frac{\partial E}{\partial \alpha} = 2\sum_{i=1}^{n}(\alpha x_i + \beta - y_i)x_i = 0, \quad \frac{\partial E}{\partial \beta} = 2\sum_{i=1}^{n}(\alpha x_i + \beta - y_i) = 0.$$

Then we have

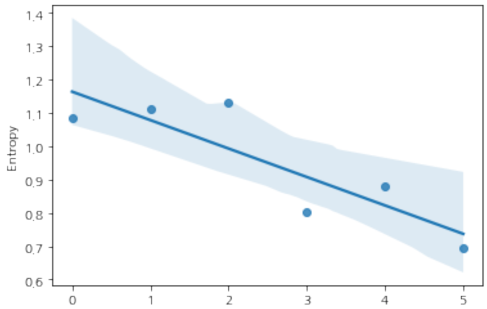
$$\begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}.$$



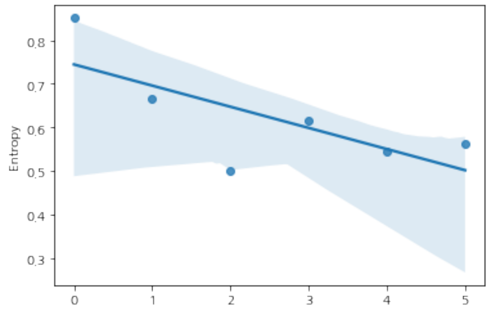
(a) Magnesium oxide



(b) Blank mask



(c) Radio telephone



(d) Entertainment robot

FIGURE 4. Entropy trend lines for selected data

Then the optimal parameters $(\hat{\alpha}, \hat{\beta})$ are

$$\hat{\alpha} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad \hat{\beta} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.$$

In our case, we set x_i as integers $0, 1, \ldots, 5$ and set y_i as the observations since our data is the type of time series. Figure 4 shows the entropy trend lines for some selected items. We can observe that the trend of entropy of magnesium oxide is increasing, and that of other items is decreasing. Through this, it can be interpreted that the import dependence on magnesium oxide is improving. Next, we describe entropy. Entropy is a concept born of thermodynamics, first introduced in 1865 by German physicist Rudolf Julius Emanuel Clausius. Entropy is applied not only in thermodynamics but also in various places such as statistical physics, economics, sociology, and information theory [5]. The information theory was first introduced in Claude Elwood Shannon's paper "A Mathematical Theory of Communication"[10]. The amount of information is a numeric value of the information we get from a message. When digitizing information, we want the amount of information in an easy-to-get message to have a low value, and the amount of information in a hard-to-get message to have a large value. If we put p_i as the probability of obtaining any message i, we define the amount of information in the message

i as follows:

$$\text{Information}(i) = -\ln p_i.$$

If defined in this way, it has a low value when the probability is high, and a high value when the probability is low. In information theory, entropy is defined as the average value of information obtained from all messages as follows:

$$\text{Entropy} = -\sum_{i=1}^{n} p_i \ln p_i.$$

Here, n is the number of all messages that can be obtained. Spurling [7] introduced the entropy in the calculation of market concentration, and Gineviiius et al. [8] compared it with various indicators related to market concentration. In this case, n is the number of companies participating in the market for the item, and $p_i \in [0, 1]$ is the market share for i companies.

Assuming a global supply chain as a market, it can be used as a meaningful indicator to explain the import dependence in the global supply chain because it can be considered as market share.

The entropy of one item imported from a total of n countries has a value between 0 and $\ln(n)$. If you import exclusively from one country (import 100% from one country), the entropy is zero. The entropy has $\ln(n)$ as the largest value if imported at the same rate from n countries. From a mathematical point of view, the entropy has a characteristic that the value changes sensitively depending on whether a particular item is monopolized or not due to the characteristics of log function. We define $\mathbf{P} = (p_1, ..., p_n)$. Now let's look at the following example.

$$\mathbf{P} = (0.9, 0.1) \quad \Rightarrow \quad \text{Entropy} = 0.3251 \qquad \text{(A)}$$
$$\mathbf{P} = (0.9, 0.05, 0.05) \quad \Rightarrow \quad \text{Entropy} = 0.3944 \qquad \text{(B)}$$
$$\mathbf{P} = (0.9, 0.025, 0.025, 0.025, 0.025) \quad \Rightarrow \quad \text{Entropy} = 0.4637 \qquad \text{(C)}$$
$$\mathbf{P} = (0.75, 0.25) \quad \Rightarrow \quad \text{Entropy} = 0.5623 \qquad \text{(D)}$$
$$\mathbf{P} = (0.6, 0.4) \quad \Rightarrow \quad \text{Entropy} = 0.6730 \qquad \text{(E)}$$

From the above observations, we have the followings:

(A)→(D)→(E): When the ratio of the number of importing countries is lowered while the number of importing countries is fixed, the entropy increases. This indicates that the proportion of highest importer has a very great influence on import dependence.

(A)→(B)→(C): It can be seen that if the ratio of top one country is fixed and the number of importing countries is increased, the entropy increases. This means that as the number of importing countries increases, the import dependence decreases. However, it is also a dangerous item as the ratio of top one country has not changed. Therefore, if the number of importing countries increases, it is necessary to check the ratio of top one importing county.

All the cases are interpreted mathematically by the following example.

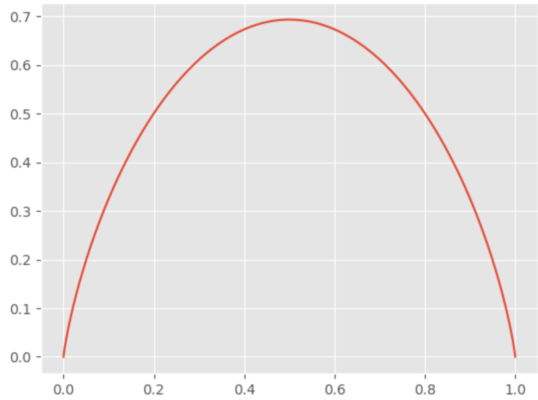


FIGURE 5. Curve of entropy

Example 3.1.

(1) *Assume that the number of importing countries is two. The import ratios for the countries are* a, b $(0 < b \leq a < 1, a + b = 1)$*, respectively. Then we have*

$$\begin{aligned} \text{Entropy} &= -a \ln a - b \ln b \\ &= -a \ln a - (1-a)\ln(1-a). \end{aligned}$$

The graph of function $g(x) = -x \ln x - (1-x)\ln(1-x)$ *is in Fig. 5. This function* $g(x)$ *changes from increasing to decreasing based on* $x = 0.5$*. Since the minimum value of* a *is 0.5, reducing* a *from 1 to 0.5 makes the entropy increase.*

(2) *Moreover, if one increases the number of importing countries with the largest ratio* a *fixed, the new import amount is generated by splitting the* b *value. Assume that it is equally divided into* n *countries. Then, we obtain that*

$$\text{Entropy} = -a \ln a - b \ln b \Rightarrow \text{Entropy} = -a \ln a - n \cdot \frac{b}{n} \ln \frac{b}{n} = -a \ln a - b \ln \frac{b}{n}.$$

Since a *and* b *are constants, the entropy after division increases as* n *increases. In other words, the entropy increases as the number of importing country increases.*

3.4.2. *The Anomaly Detection of Trend.* We selected items of interest by investigating the import size and import dependence in the previous step. Among these items of interest, items with different trends will be classified as Global supply chain risky items by comparing the recent trends in the three time series of each item's import amount, import weight, and TSI with past median trends. The trend abnormality is detected in order to reflect the industrial knowledge that risky items have trends different from before in import-related indicators in advance. Accordingly, we apply the sliding window method to group each time series into four-month units and define the anomaly score by measuring the extent to which the last 4-month data deviates from the median data. The sliding window method is known as an effective method for comparing trends in time series and detecting anomalies. Based on exploratory data analysis and prior knowledge, it is confirmed that each time series has different characteristics and is independent. The final trend anomaly score is defined by the average of each anomaly score.

We now describe the anomaly score analysis method. Suppose that the total import amount (v), import weight (w), and TSI (tsi) data of a particular item are given until October 2022. We measure how different the given 4-month data vector

$$\mathbf{v}_{202210} := [v_{202207}, v_{202208}, v_{202209}, v_{202210}]$$

is from the last 12 median data. Here, v_{202210} means the import amount on October 2022 of the item. That is, we calculate an anomaly score s_v to be described later by measuring the extent to which the vector $\mathbf{v}_{202210}$ deviates from the median matrix

$$\bar{V}_{202210} = \begin{bmatrix} \bar{v}_{202107} & \bar{v}_{202108} & \bar{v}_{202109} & \bar{v}_{202110} \\ \bar{v}_{202108} & \bar{v}_{202109} & \bar{v}_{202110} & \bar{v}_{202111} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{v}_{202206} & \bar{v}_{202207} & \bar{v}_{202208} & \bar{v}_{202209} \end{bmatrix}$$

where $\bar{v}_{202209}$ refers to the median of total import amount in September for three years, including the target year, i.e., $\bar{v}_{202209} = \mathrm{median}(v_{202209}, v_{202109}, v_{202009})$. The reason why the median values are used instead of the actual values when constructing the data matrix to be compared is to avoid the masking problem that the abnormality on the month is not properly detected due to abnormal trends that have occurred in the past. In order to effectively detect abnormal trends, the median matrix comprises median values of 3 years. The reason why the past 12 data were used is to include the same month period of the previous year of the sample to measure trend abnormality. In Fig. 6 and Fig. 7, the blue and orange lines mean actual values and 3-year median values, respectively. We can verify that the median line is smoother than the actual line. By applying the above method to each of the three time series, we compute three anomaly scores s_v, s_w, and s_{tsi} to define the trend anomaly scores as follows:

$$s := \text{mean}\{s_v, s_w, s_{tsi}\}.$$

Now, we explain how to compare the 4-month data vector $\mathbf{v}_{202210}$ with the median matrix $\bar{V}_{202210}$. It is assumed that the median matrix is a normal data to learn the past median data, which is a comparison target. Therefore, we embodied the problem with One Class Classification (OCC), a semi-supervised learning that learns only using normal data and determines whether the sample is normal. We propose Random Cut Forest(RCF), one of the methods for detecting OCC anomalies. RCF is introduced in the paper 'Robust Random Cut Forest Based Anomaly Detection On Streams' (2016) [9] and is a method that transforms the Isolation Forest algorithm (IF) (2008) [6] to match anomaly detection in streaming data. The RCF is a commercial algorithm installed in Amazon AWS Sagemaker that measures abnormal scores using Collusive Displacement (CODISP) to avoid masking problems. The RCF, which can dynamically transform trees, is known to be effective in detecting time series data anomalies using sliding window methods that bind time series data in specific time periods.

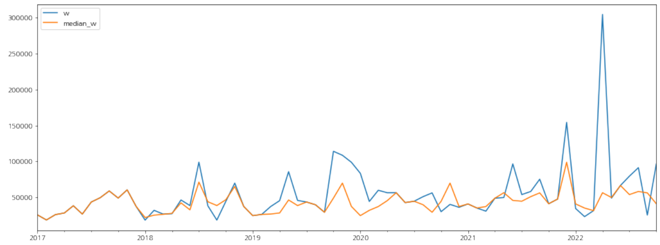


FIGURE 6. Magnetic sensor

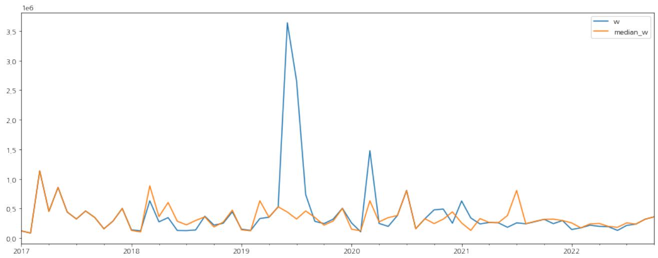


FIGURE 7. CO_2 Incubator

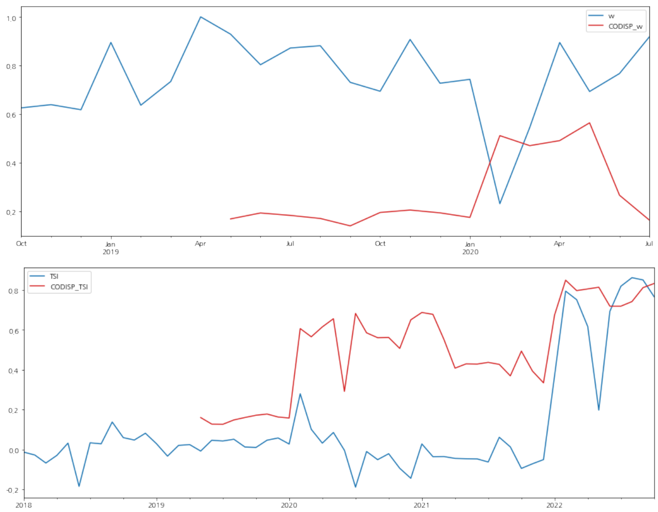


FIGURE 8. Time series data and anomaly scores for the Photodiode_illuminance sensor

Figure 8 and 9 show the time series data and anomaly scores. The blue and red lines mean an actual value (v, w, or tsi) and an anomaly score, respectively. A common feature of IF and RCF is a random tree-based bagging ensemble model that learns a large number of random trees

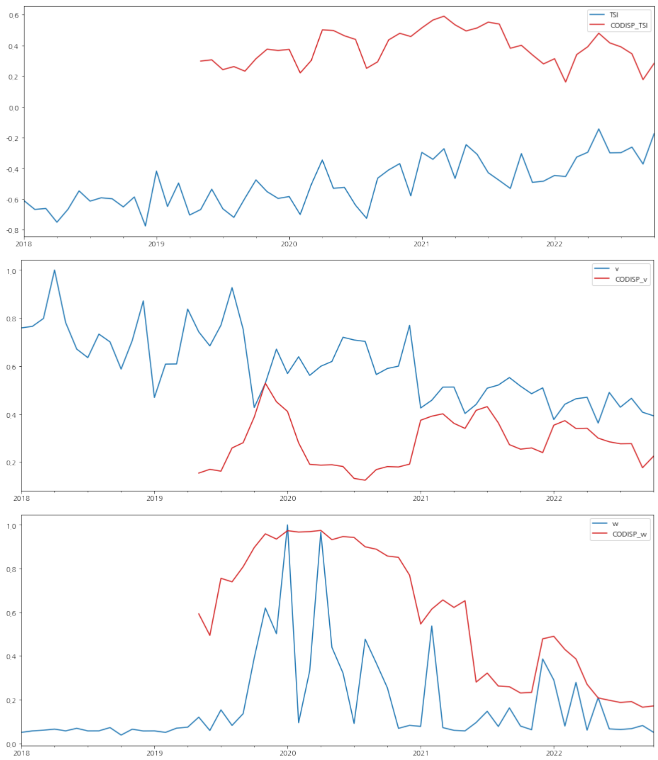


FIGURE 9. Time series data and anomaly scores for the Blank mask

(Default = 100) and utilizes them for anomaly detection. The IF shows a high AUROC value (threshold) for the Outlier Detection problem compared to the traditional anomaly detection model. Here, the AUROC value refers to a criterion for determining sensitivity or specificity. Compared to the existing distance-based and density-based anomaly detection algorithms, this shows a fast speed in anomaly detection. Since the Random Sampling method is used, it can be effectively applied to very large data. Even if there are unnecessary characteristics for anomaly detection, it is relatively not affected much.

However, IF, which shows strength in the Outlier Detection problem, is an inductive inference method that is not effective for OCC problems where only normal data exists in the learning data, and is not suitable for streaming data anomaly detection, such as assuming that all features are independent. In addition, since there are only 12 learning data, the Auto-encoder-based Reconstruction deep learning or Recurrent Neural Networks method was not considered because good results could not be guaranteed with these methods.

4. Results and Conclusion

4.1. **Results.** In this section, we describe the results of identifying global supply chain risky items based on the proposed method.

The client company provided data on export/import amount ($) and weight (kg) of global supply items from January 2017 to October 2022. Data preprocessing was performed to remove items with the same name or value out of a total of 1,772 items. In addition, only items commonly included in export data and import data were selected, and 888 items were finally selected as analysis targets. Following the client company's comment that no value in the data means no trade, it was filled with a small positive number(10^{-8}) instead of zero to prevent computational errors.

Of the 888 items, only influential items were selected from the global supply chain, excluding items with imports of less than $1 million. Among them, items with more than 90% import ratio, or those with import ratio between 75% and 90% satisfying that the slope of entropy trend line in the last 6 months was less than for equal to -0.05, were selected as items of interest. Figure 10 is the histogram for the selected items of interest. For the selected 67 items of interest,

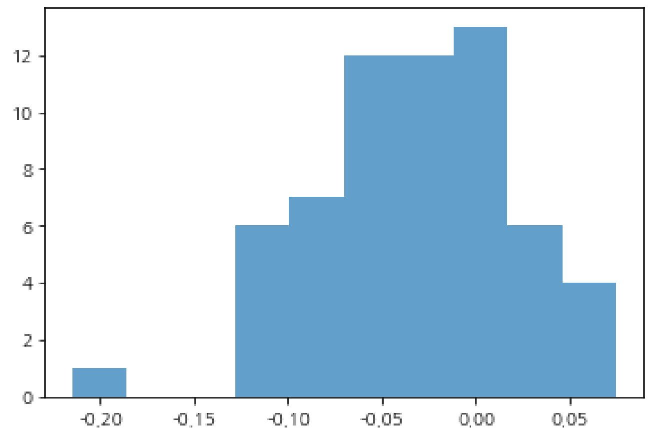


FIGURE 10. Histogram of the slope of entropy for items of interest

we evaluate the anomaly score. As described in the previous section, we use the average of CODISP score for import amount, import weight, and TSI as the anomaly score. Table 2 shows the top 10 risky items based on the anomaly score. Figure 11 shows the time series data with a high anomaly score. We can confirm that the selected items behave differently from before 2022.

4.2. **Conclusion.** We propose a novel methodology for identifying risky items based on global supply chain export/import data. This methodology selects items of interest using import size and entropy, measures the trend anomaly score of data, and selects risky items based on scores. This methodology was designed by defining the requirements through relevant prior research

Index	Risky items	CODISP
1	Photodiode_illuminance sensor	0.881667
2	Photodiode_lidar	0.870241
3	Dry etcher	0.761190
4	Evaporation	0.751125
5	4G video transceiver	0.740375
6	Wet stripper	0.733162
7	Tin oxide	0.729222
8	Lithium polymer battery	0.697667
9	Excimer laser annealing, ELA	0.642940
10	Rare earth permanent magnet	0.633065

TABLE 2. CODISP table for top 10 items

FIGURE 11. Actual data graphs for three selected risky items with a high anomaly score: Photodiode_illuminance sensor; Photodiode_lidar; Tin oxide.

and discussions with experts, and the methodology was created using data provided by the client company. In addition, it was confirmed that data with different trends were well identified by applying this methodology to the actual data.

Through this methodology, it is expected that countries and companies will be able to identify global supply risky items in advance and prepare for new normal situations: economic situations caused by infectious diseases such as the COVID-19 pandemic; and export/import regulations due to geopolitical problems. The client company will use the proposed method to identify the risky items. In addition, research topics on problems related to the global supply chain and the detection of abnormalities in time series data will be studied in the future.

ACKNOWLEDGMENTS

This work was supported by National Institute for Mathematical Sciences(NIMS) grant funded by the Korean government(MSIT) (No.NIMS-B22810000). This work was supported by the Technology Innovation Program (20020724, Building Statistics DB of Material-Parts-Equipment Industry And Global Value Chain Analysis) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea).

REFERENCES

[1] Deloitte Clients & Industries, *Deloitte Insights*, Deloitte, **22** (2022).

[2] Korea Association of Machinery Industry, *GVC Value Chain Risky Item Report*, Global Value Chain Report, Korea Association of Machinery Industry, 2022.

[3] Eunji Min, Sunkyeong Lee, *Analysis of Vulnerabilities in Korea's Import Supply Chain*, Articles in Monthly Bulletin, Bank of Korea (2022)

[4] Bawoo Kim, Yunsoo Kim, Kye Hwan Kim, *Analysis of Vulnerabilities and Impact Pathways in Korea's Industrial Supply Chain*, i-KIET Issues & Analysis, Korea Institute for Industrial Economics & Trade (2021)

[5] Alfred Wehrl, *General properties of entropy*, Reviews of Modern Physics, **50** (1978), 221.

[6] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou, *Isolation forest*, 2008 eighth ieee international conference on data mining. IEEE, 2008.

[7] D. Spuretling, *Zur Messung der wirtschafi chen Konzentration und des Wettbewerbsgrades über einige neuere Masse der Wirtschafl ichen Konzentration*, Zeitschrift für angewewandte Konjukturforschung, **16** (1970), 238-251.

[8] Romualdas Ginevičius, and Stasys Čirba, *Additive measurement of market concentration*, Journal of Business Economics and Management, **10** (2009), 191-198.

[9] Sudipto Guha, et al., *Robust random cut forest based anomaly detection on streams*, International conference on machine learning, PMLR, 2016.

[10] Claude E. Shannon, *A mathematical theory of communication*, The Bell system technical journal, **27** (1948), 379-423.