

Article

Using a Genetic Algorithm to Build a Volume Weighted Average Price Model in a Stock Market

Seung Hwan Jeong ¹, Hee Soo Lee ², Hyun Nam ³ and Kyong Joo Oh ^{1,*}¹ Department of Industrial Engineering, Yonsei University, Seoul 03722, Korea; jsh0331@yonsei.ac.kr² Department of Business Administration, Sejong University, Seoul 05006, Korea; heesoo@sejong.ac.kr³ Department of Investment Information Engineering, Yonsei University, Seoul 03722, Korea; mcgyver3@gmail.com

* Correspondence: johanoh@yonsei.ac.kr; Tel.: +82-2-2123-5720

Abstract: Research on stock market prediction has been actively conducted over time. Pertaining to investment, stock prices and trading volume are important indicators. While extensive research on stocks has focused on predicting stock prices, not much focus has been applied to predicting trading volume. The extensive trading volume by large institutions, such as pension funds, has a great impact on the market liquidity. To reduce the impact on the stock market, it is essential for large institutions to correctly predict the intraday trading volume using the volume weighted average price (VWAP) method. In this study, we predict the intraday trading volume using various methods to properly conduct VWAP trading. With the trading volume data of the Korean stock price index 200 (KOSPI 200) futures index from December 2006 to September 2020, we predicted the trading volume using dynamic time warping (DTW) and a genetic algorithm (GA). The empirical results show that the model using the simple average of the trading volume during the optimal period constructed by GA achieved the best performance. As a result of this study, we expect that large institutions will perform more appropriate VWAP trading in a sustainable manner, leading the stock market to be revitalized by enhanced liquidity. In this sense, the model proposed in this paper would contribute to creating efficient stock markets and help to achieve sustainable economic growth.

Keywords: dynamic time warping; genetic algorithm; sliding window; volume forecasting; volume weighted average price



Citation: Jeong, S.H.; Lee, H.S.; Nam, H.; Oh, K.J. Using a Genetic Algorithm to Build a Volume Weighted Average Price Model in a Stock Market. *Sustainability* **2021**, *13*, 1011. <https://doi.org/10.3390/su13031011>

Received: 29 December 2020

Accepted: 14 January 2021

Published: 20 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over time, studies to predict prices in the stock market have actively been conducted. Various studies have been carried out to predict stock prices using the auto-regressive integrated moving average (ARIMA) model, which is a time series data prediction method [1,2], as well as other methods. Pai and Lin [3] applied a hybrid methodology combining the ARIMA model and support vector machine (SVM) for stock price prediction. Wang and Leu [4] proposed a model that predicts the price trend of Taiwan's stock market by combining the ARIMA model and a neural network. Adebisi et al. [5] compared the performance of the ARIMA model and the artificial neural network (ANN) model using the stock data of the New York Stock Exchange (NYSE). Oh and Kim [6] proposed a piecewise nonlinear model using ANN to predict the stock market, which uses backpropagation neural networks to find points of continuous change in time series data. Similarly, there has been much research on applying neural network methodology to stock price prediction. Yoon and Swales [7] proved that neural networks are effective for solving complex problems such as stock price prediction, and Kohara et al. [8] showed that the neural network and prior knowledge of prediction were effective in predicting stock prices. Tsai and Wang [9] showed that when a stock price prediction model was created by combining ANN and decision tree, it showed higher accuracy than a single model. Hadavandi et al. [10] proposed a stock price prediction expert system by combining ANN and a genetic fuzzy system.

Chen et al. [11] proved that a fuzzy time series model based on the Fibonacci sequence is effective in predicting the Taiwan semiconductor manufacturing company (TSMC) stock price data and Taiwan capitalization weighted stock index (TAIEX) data. Cheng et al. [12] proposed a hybrid model for stock price prediction based on genetic algorithms and rough sets theory.

In addition to predicting stock prices of various markets, investors are also interested in stock trading volume. Trading volume is an important indicator for investors to buy or sell certain stocks. A number of studies confirmed that the trading volume has a positive correlation with the volatility of the price [13,14], and various studies predicted the price volatility based on this positive correlation by utilizing the trading volume [15,16]. Tsang and Chong [17] presented a strategy to obtain investment returns using volume-based on-balance volume (OBV) indicators, and Nedunchezian [18] conducted a study to predict the price movement of multi commodity exchange (MCX) energy using OBV indicators. However, research that predicts the actual intraday trading volume, rather than just using the trading volume as an indicator for prediction, has not been actively conducted [19]. Without sophisticated research on the intraday trading volume, large institutions are still taking the strategy to consume liquidity using the simple average of the trading volume in the past period or to consume liquidity in the market at the beginning or end of the market when the trading volume is high.

It is widely known that the extensive trading volume by large institutions has a significant impact on the market liquidity and the volume weighted average price (VWAP) model is used by large institutions. Based on a distribution of intraday trading volume in a stock market, the VWAP model allocates trades in a way that reduces the impact of large institutions' trading on the stock market liquidity. Thus, it is critical to predict the trading volume accurately for VWAP trading. In this study, we propose optimal models to predict stock trading volume using dynamic time warping (DTW) and a genetic algorithm (GA). DTW and GA have been widely used for developing various investment strategies. Previous studies proposed a pattern matching trading system using DTW to predict exchange rates and stock prices [20–23]. Additionally, GA has been used for predicting stock indices, real estate auction prices and appraisals, and was also used to optimize IPO investment strategies or trading strategies that hedge options [24–29].

In the empirical study, we predicted the trading volume using DTW and GA using the trading volume data of the Korean stock price index 200 (KOSPI 200) futures index from December 2006 to September 2020. We used four methods to predict the trading volume and compare their performance. Those methods include calculating a simple moving average of trading volume over the past 20 days, DTW trading volume, DTW trading volume based on grouping, and GA trading volume. We employed a GA in three ways to forecast trading volume: a fixed 20-day GA weighted average method, an optimal GA weighted dynamic period method, and a simple average of the optimal GA period method. Our empirical results show that predicting the trading volume using a simple average of the optimal GA period method achieves the best performance.

This paper is organized as follows: Section 2 presents the literature review, Section 3 presents the data and methodology used in this paper. Section 4 shows the empirical study, and Section 5 presents the conclusions of our study.

2. Literature Review

2.1. Volume Weighted Average Price

The trading of large institutions such as pension funds consumes much liquidity, which readily impacts the market. Therefore, large institutions generally try to minimize the impact on the market by dividing the order volume, which is called a careful discretionary (CD) order. Typical CD ordering methods include time-weighted average price (TWAP) and volume-weighted average price (VWAP) [30]. TWAP focuses on time and distributes the order quantity uniformly. For example, when an investor wants to buy 6000 shares of a certain stock over 6 h, buying 1000 shares per hour is called TWAP. Conversely,

VWAP focuses on the trading volume [31–33]. Generally, the intraday trading volume shows a U-shaped curve with much trading at the beginning or end of the market and relatively few trades in the middle of the market [34–36]. If a CD order focuses on trades that have less impact on the market, VWAP, which allocates relatively more trades in the middle of the market and relatively fewer trades in the early or late market, is more effective than TWAP [37]. Most of the institutions trade 50% of the transactions using VWAP [38]. It is critical to predict the trading volume accurately when VWAP is used. A CD order customarily uses a simple average volume over the past 20 days to determine VWAP trading.

2.2. Dynamic Time Warping

DTW is an algorithm that measures the similarity between two different patterns. Typically, a simple distance measurement such as the Euclidean distance is sufficient when a distance between two sequences in alignment is measured, but when the x-axis of the two sequences is not aligned, the x-axis of the two sequences must be warped [39]. Essentially, DTW warps the x-axis to find the distance between two sequences. DTW is a pattern detection algorithm that has been primarily used in the field of speech recognition [40,41]. Alternatively, this algorithm can also be used to compare time series data of different lengths of time [42]. Keogh and Ratanamahatana [43] proved that DTW is a more powerful distance measurement method than any other method when distances between time series data are measured.

The x-axis warping of two sequences that are not aligned on the x-axis is as follows: first, we create an $m \times n$ matrix to compare the two time series $X = (x_1, x_2, \dots, x_m)$ with length m , and $Y = (y_1, y_2, \dots, y_n)$ with length n . The $m \times n$ matrix consists of the Euclidean distance between x_i in row m and y_j in column n denoted by $d(x_i, y_j)$. An optimal warping path can be found by finding the minimum path from $d(x_1, y_1)$ to $d(x_m, y_n)$. When finding the minimum path, it should be sought only in a direction that does not retreat. The formula is as follows:

$$DTW(X, Y) = \min \left(\sum_{k=1}^K d(x_{i_k}, y_{j_k}) / K \right) \text{ where } \max(m, n) \leq K < m + n - 1 \quad (1)$$

In the equation, $d(x_{i_k}, y_{j_k})$ is $\sqrt{(x_{i_k} - y_{j_k})^2}$ ($1 \leq i_k \leq m, 1 \leq j_k \leq n$), and K in the denominator is used to compensate for values with different warping path lengths.

The optimal warping path for calculating $DWT(X, Y)$ can be obtained through dynamic programming by recurring the following equation:

$$DTW(i_k, j_k) = d(i_k, j_k) + \min(DTW(i_k - 1, j_k), DTW(i_k, j_k - 1), DTW(i_k - 1, j_k - 1)) \quad (2)$$

A number of studies use DTW for time series analysis. Tsinaslanidis and Kugiumtzis [20] predicted the GBP/USD exchange rate using DTW and perceptually important points (PIP). Using DTW, Nakagawa et al. [21] found a past time series pattern similar to the present pattern based on which future stock prices are predicted. Tsinaslanidis [22] used DTW to predict bullish and bearish markets for a number of NYSE-listed stocks, and Kim et al. [23] constructed pattern matching trading system (PMTS) for KOSPI 200 futures index time series data using DTW. This PMTS was proposed as a way to determine the clearing strategy in the afternoon by utilizing the trend of the morning market as a specific pattern.

2.3. Genetic Algorithm

The genetic algorithm (GA) introduced by John Holland in the early 1970s is a probabilistic search algorithm based on the mechanics of natural selection and natural genetics [44]. The GA repeats the fitness evaluation and probabilistic selection of each chromosome over generations based on a population or a set of chromosomes. The fitness function is used to evaluate the fitness of a chromosome, and then a chromosome with a high fitness value is randomly selected. In addition, some of the selected chromosomes are subjected to crossover and mutation to create new chromosomes, which are called the next

generation of chromosomes. The basic concept of a GA is to obtain a final chromosome with a high average fitness value through this process.

GAs incorporating evolutionary algorithms such as selection, crossover, mutation, and inheritance have been widely used to find optimal solutions to complex problems in various fields. In particular, they have been widely used to find the optimal solution for random data in financial markets. Allen and Karjalainen [24] used a GA to learn the optimal transaction rules for the S&P 500 index, and Kim and Han [25] proposed a hybrid model combining a GA and an artificial neural network to predict the stock index. Ahn et al. [26] proposed an effective model for predicting a real estate appraisal by combining ridge regression with a GA, and Kang et al. [27] proved that a GA is effective in predicting real estate auction prices. Song et al. [28] proposed an option hedging system using a GA to improve the option hedging effect, and Kim et al. [29] proposed a machine learning investment strategy for Korean IPO stocks utilizing a rough set and GA.

3. Materials and Methodology

Our prediction analysis was conducted using four major steps, as shown in Figure 1, which was the basis of our analysis. Using the KOSPI 200 futures index trading volume from December 2006 to September 2020, we first calculated a simple moving average of 20 days to use it as a benchmark for six optimal models proposed in this study. In the next step, we used a DTW pattern matching model to predict trading volume. Then we proposed the use of grouping for the DTW pattern matching model. Groupings based on KOSPI 200 futures volume (volume group A) and KOSPI 200 futures index (index group B) were constructed to classify the current trading volume trend. In the last step, we proposed three GA models to predict the optimal trading volume: a fixed 20-day GA weighted average method, an optimal GA weighted dynamic period method, and a simple average of the optimal GA period method.

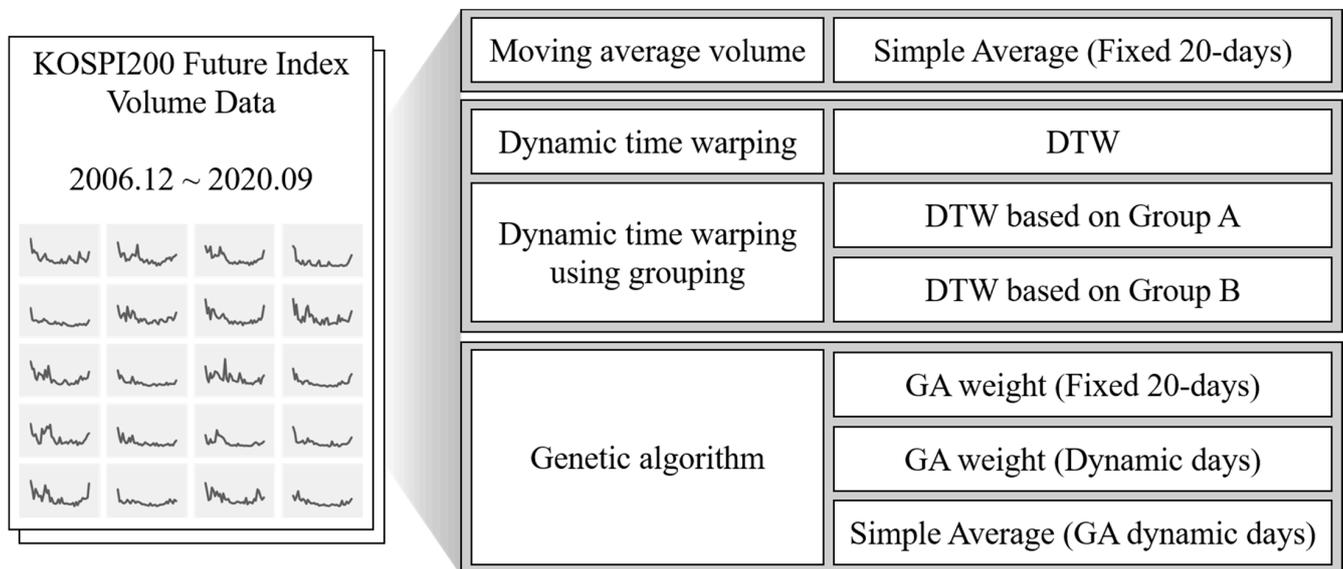


Figure 1. Structure diagram of data and methodology.

3.1. Data Collection and Preprocessing

For the empirical study, we used the minute trading volume data of the KOSPI 200 futures index from 1 December 2006 to 29 September 2020. This study used futures market data but ultimately incorporated it into the spot market. The KOSPI 200 futures index is a product based on the KOSPI 200, which is calculated with the market capitalization of 200 stocks listed on the South Korean securities market. The settlement months of the futures contract are on the second Thursday of March, June, September, and December.

The transaction unit is obtained by multiplying the KOSPI 200 futures price by KRW 250,000, and the price is expressed as a number (points) of the KOSPI 200 futures. The regular trading hours of the Korean derivatives market are 9:00–15:45 (405 min), and a total of 1,268,280 min of trading volume data over the sample period were used for this empirical study. Considering the different trading hours of the Korean derivatives market until July 2016, we used a 45-month test period from January 2017 to September 2020 in order to maintain the consistency of the trading volume. Additionally, since the total trading volume and minute trading volume differed from day to day, a scaling process was performed to divide the minute trading volume by the total trading volume of the corresponding date. Consequently, while maintaining the shape of the intraday trading volume, the problem of unit differences in the daily trading volume was solved.

In the DTW and GA optimization empirical study, the test data was used from January 2017 to September 2020. For the DTW training data, the trading volume from December 2006 to the day before the test day, and in the GA optimization empirical study, the sliding window method was employed to set the training and test periods. Figure 2 shows the structure of the sliding window method. The sliding window method first divides each window into training and test periods. Then, it repeats the training and test by sliding the training and the test periods as shown in Figure 2 [23,45,46].

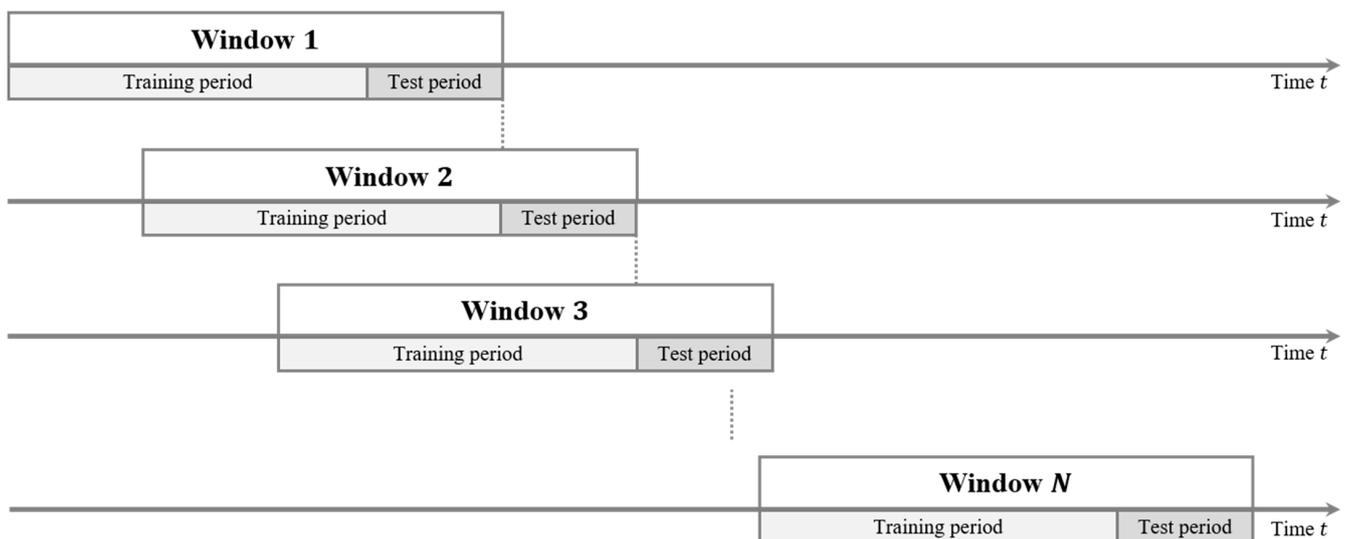


Figure 2. Structure of the sliding window method.

3.2. Trading Volume Forecasting Methodology

3.2.1. Moving Average Volume

The simple average trading volume, which is commonly used as a baseline model, is simply averaged over the past 20 days and then used as a forecast of tomorrow's trading volume. For example, the value of the trading volume at 9:01 on day t can be predicted by simply averaging the value of the trading volume at 9:01 from day $t - 1$ to day $t - 20$. The formula for the moving average volume is as follows:

$$\hat{V}_{t+1} = \frac{V_t + V_{t-1} + \dots + V_{t-20}}{20} \quad (3)$$

In the above formula, \hat{V}_{t+1} means the predicted trading volume on the day $t + 1$, and V_t means the actual trading volume on the day t .

3.2.2. Dynamic Time Warping-Based Trading Volume

In DTW experiments, reference data is required to be compared with the training data. After setting the reference period, we calculated the DTW value with the reference data for each training data and made a decision. In this study, reference data was organized in ascending order of DTW values based on the shape of the trading volume on day $t - 1$. Then, the days in which the DTW value is in the top 1, top 1%, top 5%, and top 20 were selected. Last, the trading volume on the next day after the selected day was used as the predicted trading volume on day t .

3.2.3. Dynamic Time Warping-Based Trading Volume Using Grouping

In the DTW experiment, the concept of grouping was additionally introduced to reduce the diversity of stock market patterns. The grouping method was used as a filter because it is difficult to recognize the trend of the trading volume simply by comparing the DTW values. Group A was constructed to classify the current trading volume trend by analysing the trend of the volatility index of KOSPI 200 (VKOSPI) and the moving average (MA) of the trading volume of the KOSPI 200 futures index.

In Figure 3, group A is a combination of the VKOSPI group and KOSPI 200 futures volume group. If the VKOSPI 5-day MA was lower than the VKOSPI 20-day MA, 60-day MA, and 120-day MA, the VKOSPI group was set to 1. In addition, if the VKOSPI 5-day MA was higher than other MAs, the VKOSPI group was set to 3. In other cases, the VKOSPI group was set to 2. In a similar manner, the KOSPI200 futures volume group was classified into three groups based on volume MA. Finally, nine group As were formed by combining the VKOSPI group and KOSPI 200 futures volume group.

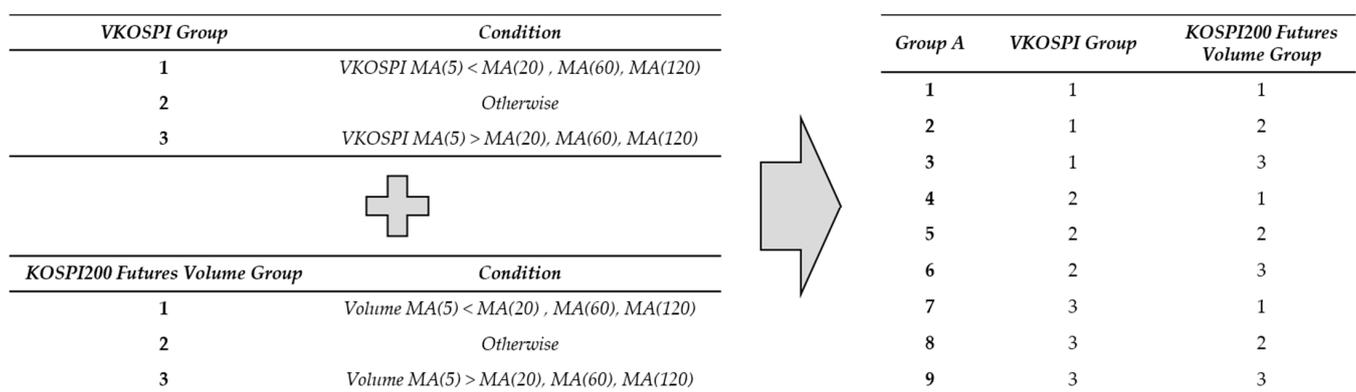


Figure 3. Group A combining the VKOSPI group and KOSPI 200 futures volume group.

Group B is similar to group A, but it was constructed to classify the current trading volume trend by the MA of the KOSPI 200 futures index itself, not the trading volume of the KOSPI 200 futures index. As shown in Figure 4, group B is a combination of the VKOSPI group and KOSPI 200 futures index group. In the case of the KOSPI 200 futures index group, if the KOSPI 200 future index 5-day MA was lower than the KOSPI 200 futures index 20-day MA, 60-day MA, and 120-day MA, the KOSPI 200 futures index group was set to 1. In addition, if the KOSPI 200 future index 5-day MA was higher than other MAs, the KOSPI 200 futures index group was set to 3. In other cases, the KOSPI 200 futures index group was set to 2. Finally, nine group Bs were formed by combining the VKOSPI group and KOSPI 200 futures index group.

VKOSPI Group	Condition
1	VKOSPI MA(5) < MA(20), MA(60), MA(120)
2	Otherwise
3	VKOSPI MA(5) > MA(20), MA(60), MA(120)

KOSPI200 Futures Index Group	Condition
1	Volume MA(5) < MA(20), MA(60), MA(120)
2	Otherwise
3	Volume MA(5) > MA(20), MA(60), MA(120)

Group A	VKOSPI Group	KOSPI200 Futures Index Group
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

Figure 4. Group B combining the VKOSPI group and KOSPI 200 futures index group.

After adding group A or group B as a filtering condition, the DTW values were arranged in ascending order based on the reference data of the same group. As described in Section 3.2.2, the days in which the DTW value is in the top 1, top 1%, top 5%, and top 20 were selected as reference data. Then, the trading volume on the next day after the selected day was used as the forecast of the trading volume on day t .

3.2.4. Genetic Algorithm-Based Trading Volume

The GA experiment was conducted in three ways. The first GA experiment used the trading volume from day $t - 1$ to day $t - 20$, as described in Section 3.2.1, but applied a weighted average using GA optimal weights, not simple averages. This is an optimization method to obtain better results than the simple average of 20 days, which has been widely used. After calculating the weight ($W_{t-1}, W_{t-2}, \dots, W_{t-20}$) that best predicts the trading volume of day t based on the minute-by-minute trading volume from day $t - 1$ to day $t - 20$, these weights were used to predict the trading volume of day $t + 1$. The formula for predicting the trading volume on day $t + 1$ (\hat{V}_{t+1}) is as follows:

$$\hat{V}_{t+1} = V_t \times W_{t-1} + V_{t-1} \times W_{t-2} + \dots + V_{t-19} \times W_{t-20} \quad (4)$$

In the second GA experiment, we optimized the past period to be used to predict the trading volume on day t , rather than fixing the trading volume from day $t - 1$ to day $t - 20$. The period N for period optimization was set to a value between 20 and 60 days (or one to three months on a business day basis). After determining the N value and weight ($W_{t-1}, W_{t-2}, \dots, W_{t-N}$) that minimize the mean absolute percentage error (MAPE) between the volume (V_t) and the predicted volume on day t (\hat{V}_t), the trading volume on day $t + 1$ (\hat{V}_{t+1}) was predicted. The formula for predicting the trading volume on days t (\hat{V}_t) and $t + 1$ (\hat{V}_{t+1}) is as follows:

$$\hat{V}_t = V_{t-1} \times W_{t-1} + V_{t-2} \times W_{t-2} + \dots + V_{t-N} \times W_{t-N} \quad (5)$$

$$\hat{V}_{t+1} = V_t \times W_{t-1} + V_{t-1} \times W_{t-2} + \dots + V_{t-N-1} \times W_{t-N} \quad (6)$$

In the last GA experiment, similar to the second GA experiment, after optimizing the period to be used for model training, a simple average of the data for the period was used to predict the volume. First, we found the GA weight and period N that showed the lowest MAPE and weighted average using the GA optimal weight to predict the trading volume on day t . The period N was selected from the period between the past 20 days ($t - 1, t - 2, \dots, t - 20$) and 60 days ($t - 1, t - 2, \dots, t - 60$). Then, only the period N was used to predict the trading volume on day $t + 1$, without the GA weight. The trading volume from day t to day $t - N + 1$ was simply averaged and used as the predicted

volume on day $t + 1$. The formula for predicting the trading volume on days t and $t + 1$ is as follows:

$$\hat{V}_t = V_{t-1} \times W_{t-1} + V_{t-2} \times W_{t-2} + \dots + V_{t-N} \times W_{t-N} \quad (7)$$

$$\hat{V}_{t+1} = \frac{V_t + V_{t-1} + \dots + V_{t-N-1}}{N} \quad (8)$$

3.3. Performance Measure

As a performance measure, MAPE, which is an index representing the extent to which the error accounts for the actual value, was used. MAPE is the average of the extent to which the difference between the actual value and the predicted value occupies the actual value for all data. The formula for MAPE on day t is as follows:

$$MAPE_t = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (9)$$

In the above equation, Y_i is the actual minute trading volume, \hat{Y}_i is the predicted minute trading volume, and n is the total number of trading volume data per day.

4. Empirical Study

For the empirical study on the various methods presented in Section 3, the test period was set from January 2017 to September 2020. First, we calculated a simple average trading volume over a commonly used 20-day period introduced in Section 3.2.1. The average of the minute trading volume from day $t - 1$ to day $t - 20$ was calculated to test the trading volume on day t (\hat{V}_t). Then, we compared the predicted trading volume using a simple average and the actual trading volume on day t (V_t) by the MAPE.

Figure 5 shows the daily MAPE in Equation (9) between the predicted volume using a 20-days simple average and the actual volume. As shown in Table 1, the average value of the MAPE within the test period of the predicted trading volume is 0.3845 and the variance of the MAPE is 0.01692.

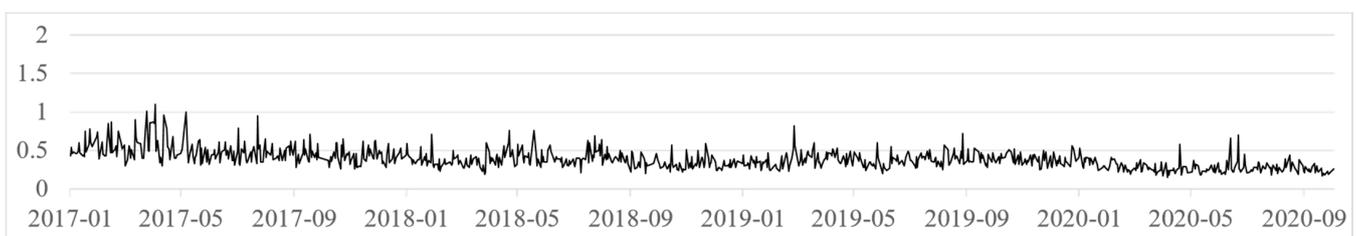


Figure 5. Daily MAPE between the predicted volume using a simple average and the actual volume.

Table 1. Average and variance of daily MAPE in the test period for 20-days simple average.

Model	Average	Variance
20-days Simple average	0.3845	0.01692

Next, for the DTW empirical study proposed in Section 3.2.2, the reference data period was set from December 2006 to the day before the test day. The DTW values were calculated for all reference data for day t , and then DTW values were sorted in ascending order. Among the DTW values sorted in ascending order, the trading volume of the next day after the day corresponding to the top 1 ranking (DTW Top 1) was used as the forecast of the trading volume on day $t + 1$ day (\hat{V}_{t+1}). In a similar manner, the trading volume of the next day after the day corresponding to the top 1% ranking (DTW 1%), the top 5% ranking (DTW 5%), and the top 20 ranking (DTW Top 20) was used as a forecast of the trading volume on day $t + 1$. In addition, we calculated the MAPE between the predicted

trading volume (\hat{V}_{t+1}) and the actual trading volume on day $t + 1$ (V_{t+1}). Figure 6 shows the daily MAPE between the predicted volume using DTW and the actual volume.

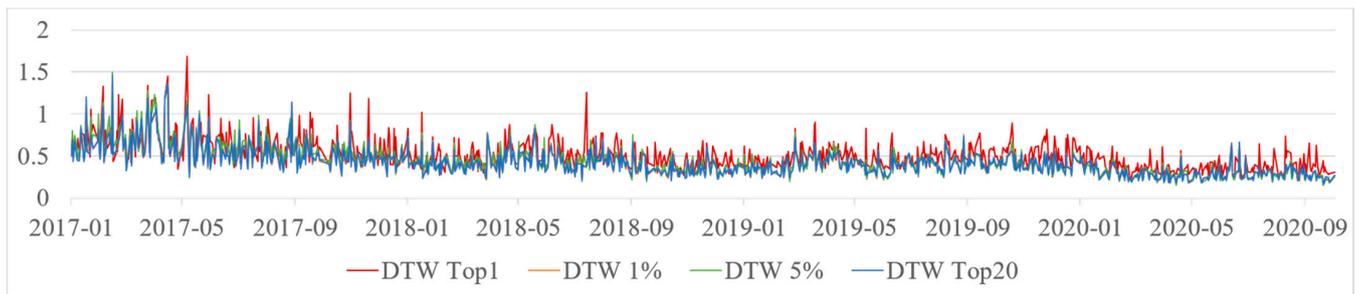


Figure 6. Daily MAPE between the predicted volume using DTW and the actual volume.

As displayed in Figure 6, DTW Top 1 generally shows a relatively higher MAPE value than other DTW methods. As shown in Table 2, the average MAPE for the test period was 0.5166 for DTW Top 1, 0.4226 for DTW 1%, 0.4414 for DTW 5%, and 0.4226 for DTW Top 20. The variance of MAPE for the test period was 0.03392 for DTW Top 1, 0.02768 for DTW 1%, 0.03252 for DTW 5%, and 0.02740 for DTW Top 20. When only DTW is used, it is best to use the volume of the day following the Top 20 as the predicted volume. However, the DTW Top 20 method underperformed the 20-day simple average trading volume method.

Table 2. Average and variance of daily MAPE in the test period for the DTW model.

Model	Average	Variance
DTW Top1	0.5166	0.03392
DTW 1%	0.4226	0.02768
DTW 5%	0.4414	0.03252
DTW Top20	0.4226	0.02740

Next, we conducted an experiment that adds the concept of grouping to the DTW method as proposed in Section 3.2.3. The DTW based on the grouping method is an experiment in which only the grouping filter rule is added to the DTW experiment. Group A was set as shown in Figure 3. First, in order to conduct a DTW experiment based on group A, group A reference data was formed by collecting dates corresponding to the same group as day t from the reference data. In addition, based on the group A reference data, the DTW experiment was carried out in the same way as before.

Figure 7 shows the daily MAPE between the predicted volume using DTW based on group A and the actual volume. As illustrated in Figure 7 and Table 3, the DTW experiment based on group A also showed the worst performance in the case of DTW Top 1. As shown in Table 3, the average MAPE for the test period was 0.5193 for DTW (group A) Top 1, 0.4501 for DTW (group A) 1%, 0.4447 for DTW (group A) 5%, and 0.4470 for DTW (group A) Top 20. The variance of MAPE for the test period was 0.03896 for DTW (group A) Top 1, 0.03078 for DTW (group A) 1%, 0.03469 for DTW (group A) 5%, and 0.03491 for DTW (group A) Top 20. DTW based on group A showed overall lower performance than the simple DTW result.

In the case of DTW based on group B, the experimental process was the same as DTW based on group A, but group B was set as shown in Figure 4. Figure 8 displays the daily MAPE between the predicted volume using DTW based on group B and the actual volume. As shown in Table 4, the results of DTW (group B) Top 1 was also the worst. The average MAPE for the test period was 0.5154 for DTW (group B) Top 1, 0.4493 for DTW (group B) 1%, 0.4389 for DTW (group B) 5%, and 0.4425 for DTW (group B) Top 20. The variance of MAPE for the test period was 0.03461 for DTW (group B) Top 1, 0.03054 for DTW (group B) 1%, 0.03239 for DTW (group B) 5%, and 0.03398 for DTW (group B) Top 20. Subsequently,

all of the DTW-related experiments exhibited lower performance than the 20-day simple average trading volume method.

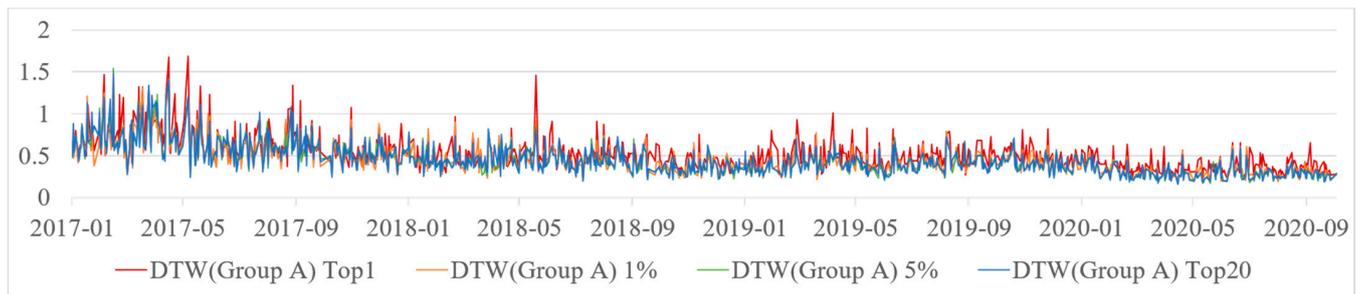


Figure 7. Daily MAPE between the predicted volume using DTW based on group A and the actual volume.

Table 3. Average and variance of daily MAPE in the test period for the DTW model based on group A.

Model	Average	Variance
DTW(Group A) Top1	0.5193	0.03896
DTW(Group A) 1%	0.4501	0.03078
DTW(Group A) 5%	0.4447	0.03469
DTW(Group A) Top20	0.4470	0.03491

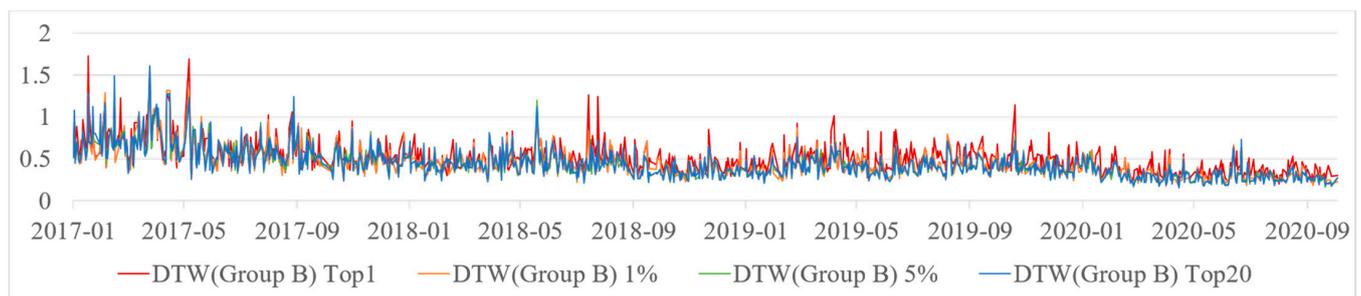


Figure 8. Daily MAPE between the predicted volume using DTW based on group B and the actual volume.

Table 4. Average and variance of daily MAPE in the test period for the DTW model based on group B.

Model	Average	Variance
DTW(Group B) Top1	0.5154	0.03461
DTW(Group B) 1%	0.4493	0.03054
DTW(Group B) 5%	0.4389	0.03239
DTW(Group B) Top20	0.4425	0.03398

Finally, an experiment was conducted using the genetic algorithm as proposed in Section 3.2.4. First, we optimized the GA weights, where a weighted average of trading volume from day $t - 1$ to day $t - 20$ was calculated to predict trading volume on day t (\hat{Y}_t). Then, the GA weight optimized volume on day t was used as the predicted volume on day $t + 1$. Figure 9 shows the daily MAPE between the predicted volume using GA and the actual volume. As shown in Table 5, the average MAPE of the fixed 20-day GA weighted average volume for the test period was 0.3892, which was higher than those of the 20-day simple average volume, but it showed better results than those from the DTW experiments.

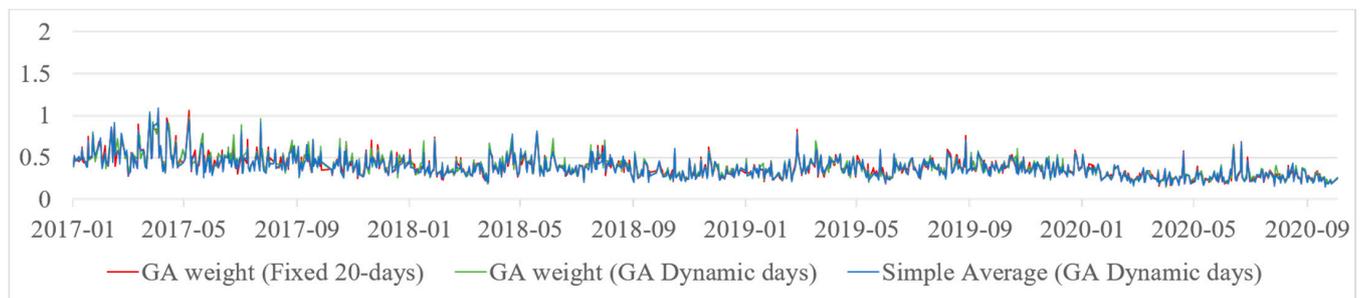


Figure 9. Daily MAPE between the predicted volume using GA and the actual volume.

Table 5. Average and variance of daily MAPE in the test period for GA model.

Model	Average	Variance
GA weight (Fixed 20-days)	0.3892	0.01754
GA weight (GA Dynamic days)	0.3938	0.01809
Simple Average (GA Dynamic days)	0.3815	0.01686

Given the better performance of the GA weighted average volume, we tried to optimize the weight and period simultaneously. After applying the GA weight to period N ($N = 20, 21, \dots, 60$) based on day t , the period N and GA weight generating the lowest MAPE were selected, and the period N and GA weight were used to predict the volume on day $t+1$. The count of windows per optimal GA period N is shown in Table 6. The average MAPE of the optimal GA weighted dynamic period volume for the test period was 0.3938, which was also higher than those of the 20-day simple average volume, but it showed better results than those from the DTW experiments.

Table 6. Count of windows per optimal GA period N .

Period N	Count of Windows	Period N	Count of Windows	Period N	Count of Windows
Period 20	28	Period 34	16	Period 48	28
Period 21	20	Period 35	45	Period 49	19
Period 22	19	Period 36	25	Period 50	10
Period 23	18	Period 37	20	Period 51	23
Period 24	11	Period 38	26	Period 52	43
Period 25	14	Period 39	19	Period 53	20
Period 26	5	Period 40	25	Period 54	27
Period 27	15	Period 41	19	Period 55	42
Period 28	18	Period 42	29	Period 56	24
Period 29	16	Period 43	19	Period 57	21
Period 30	12	Period 44	17	Period 58	32
Period 31	18	Period 45	16	Period 59	29
Period 32	9	Period 46	22	Period 60	66
Period 33	18	Period 47	15		

Therefore, we used only the optimal GA period N volume on day t , as summarized in Table 6. The optimal GA period N volume on day t was used to predict the trading volume on day $t+1$, and in this case, a simple average was used instead of the GA weighted average. As a result, the simple average of the optimal GA period N volume showed better performance than the 20-day simple average and the fixed 20-day GA weighted volume. The MAPE of the simple average of the GA dynamic period volume for the test period was 0.3815.

Last, we compared the average and the variance of MAPE in the test period for all experiments. As shown in Figure 10, the simple average of the optimal GA period volume

for the test period achieved the best performance. In addition, the GA method was found to outperform the DTW method for predicting trading volume.

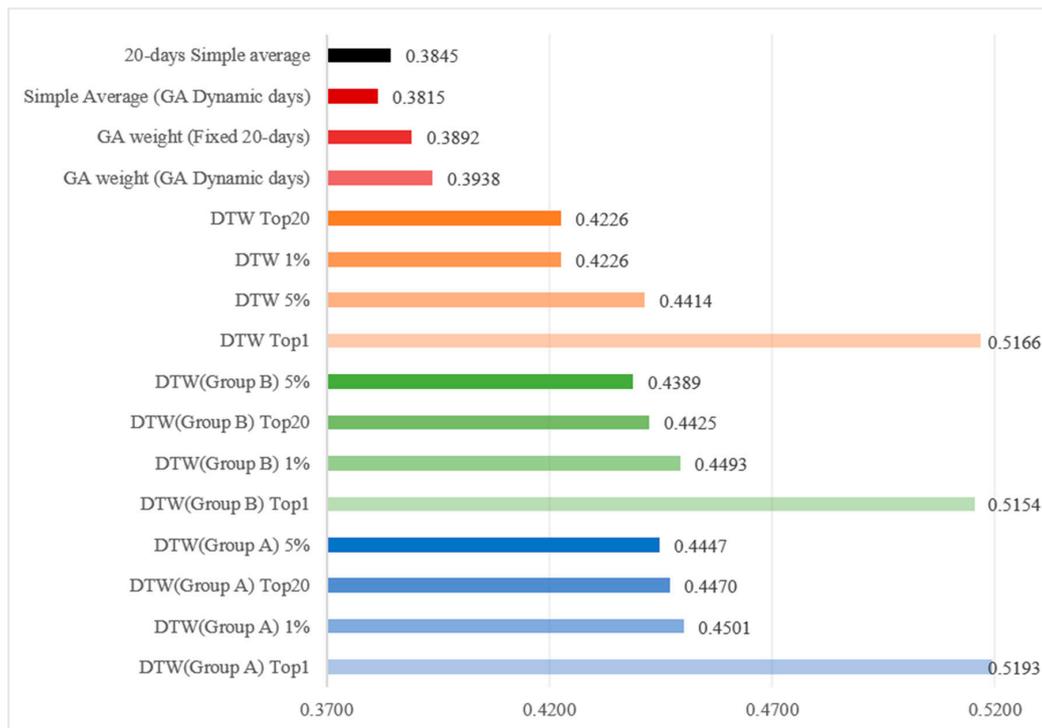


Figure 10. Average daily MAPE in the test period for all experiments.

Figure 11 illustrates that the model of the simple average of the optimal GA period volume showed the lowest variance over the test period. However, it was not found to be significantly lower than that of the model of 20-days simple average.

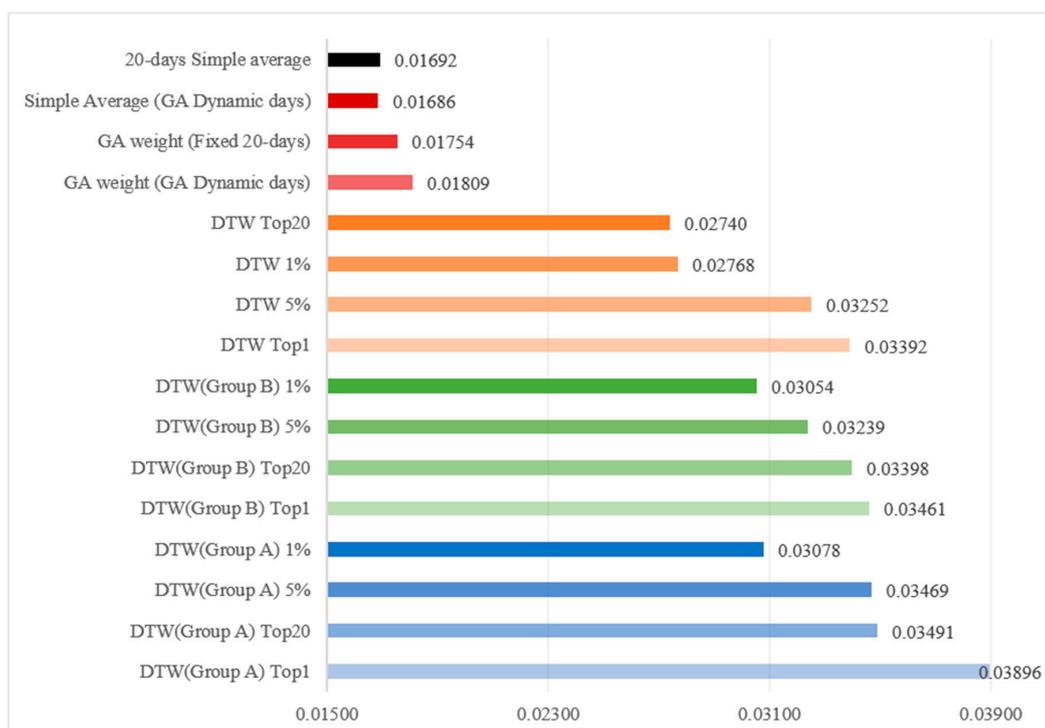


Figure 11. Variance daily MAPE in the test period for all experiments.

For more formal evaluation, we performed the paired *t*-test for MAPE to compare the predictive power of the GA models with the baseline model of 20-days simple average. Test results in Table 7 indicate that MAPE of the simple average of the optimal GA period model is significantly lower than that of the baseline model.

Table 7. Results of the paired *t*-test for MAPE between the 20-days simple average and the simple average of the optimal GA period model.

Paired <i>t</i> -Test ($\alpha = 0.05$)	20-Days Simple Average	Simple Average (GA Dynamic Days)
Mean	0.384525	0.381533
Variance	0.016915	0.016857
Number of observations	918	918
<i>t</i> statistic		4.2453
<i>p</i> -value		1.2×10^{-5}

Additionally, we performed the F-test to compare the variation in performance of the GA model and the baseline model. As shown in Table 8, variance of the simple average of the optimal GA period model was not found to be significantly lower than that of the baseline model.

Table 8. Results of the F-test for MAPE variance between the 20-days simple average and the simple average of the optimal GA period model.

F-Test ($\alpha = 0.05$)	20-Days Simple Average	Simple Average (GA Dynamic Days)
Mean	0.384525	0.381533
Variance	0.016915	0.016857
Number of observations	918	918
F statistic		1.003465
<i>p</i> -value		0.479124

5. Discussion and Concluding Remarks

In this paper, we presented four methods for predicting the trading volume in the stock market using DTW and GA: the simple moving average method, the DTW method, the DTW based on the grouping method, and the GA method. Our empirical study shows that predicting the trading volume using GA achieves the best performance.

The DTW method is known to be effective in finding trading volume similar to current trading volume. However, it has been found to have limitations in predicting future trading volume using historical data of trading volume. According to the results of the three DTW methods, classifying markets based on Group A and B did not improve predictive power. Nevertheless, when comparing DTW based on market grouping, the results of DTW with Group B combining the VKOSPI group and KOSPI 200 futures index group were found to be more effective than DTW with Group A combining the VKOSPI group and KOSPI 200 futures volume group. This result suggests that market grouping based on index value is more appropriate than that based on volume when classifying the market to predict trading volume.

We also suggest three ways to use GA to forecast trading volume: a fixed 20-day GA weighted average method, an optimal GA weighted dynamic period method, and a simple average of the optimal GA period method. The simple average of the optimal GA period method performed better than the conventional 20-day simple average method. It was expected that the GA weights would play a key role in predicting trading volume with a particular pattern, but our empirical results did not show improvement of predictive power by the GA weight method. It was found that the simple average of the GA optimal period outperformed the GA weight methods. This result implies that when predicting the trading volume, the number of periods used has a stronger impact on the predictive

power than the weight of the days for prediction. Compared with the DTW methods, the GA method generates optimal weights and periods using the latest market data, which results in higher predictive power than the DTW methods.

In the literature, research that predicts the actual intraday trading volume in a stock market, rather than just using the trading volume as an indicator for prediction of stock price or market volatility, has not been actively conducted. It is widely known that large institutions are taking a strategy to consume liquidity using the simple average of the trading volume in the past period or to consume liquidity in the market at the beginning or end of the market when the trading volume is high. Thus, in this study, we propose to use sophisticated tools for predicting trading volume in a stock market and our empirical results show slightly better predictive power of the simple average of the optimal GA period method than the baseline method for predicting trading volume in a stock market. Rather than using the 20-day simple average volume for prediction, large institutions are expected to appropriately consume liquidity and bring market vitality by applying a proposed model that optimizes the number of periods by GA. As a result of this study, we expect that large institutions will perform more appropriate VWAP trading in a sustainable manner, leading the stock market to be revitalized by enhanced liquidity. In this sense, the model proposed in this paper contributes to creating efficient stock markets and helps to achieve sustainable stock markets.

This study has potential limitations. As shown in the empirical results, the predictive power of the GA models was not improved significantly compared to the baseline method. In addition, the GA models proposed in this paper were based on the trading volume data of the KOSPI 200 futures index in Korean markets. Therefore, the empirical results are limited to Korean market data. Based on the idea of our GA models, future research can be enriched by developing a trading volume prediction model that combines various artificial intelligence methodologies and GA. In addition, various ranges of parameters can be applied to the GA models. In this study, a prediction was made by setting a day as one window. However, in the future, a more precise prediction study could be conducted by setting a minute as one window.

Author Contributions: Project administrator, K.J.O.; Proposing methodology, Validation, Writing—Original Draft Preparation, Formal analysis, S.H.J.; Programming, Resources and data curation, H.N.; Writing—Review & Editing, H.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study could be subscribed from KOSCOM's Check Expert system. Authors are restricted from sharing the data used in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ariyo, A.A.; Adewumi, A.O.; Ayo, C.K. Stock price prediction using the ARIMA model. In Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 26–28 March 2014; pp. 106–112.
2. Mondal, P.; Shit, L.; Goswami, S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *Int. J. Comput. Sci. Eng. Appl.* **2014**, *4*, 13. [\[CrossRef\]](#)
3. Pai, P.F.; Lin, C.S. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega* **2005**, *33*, 497–505. [\[CrossRef\]](#)
4. Wang, J.H.; Leu, J.Y. Stock market trend prediction using ARIMA-based neural networks. In Proceedings of the International Conference on Neural Networks, Washington, DC, USA, 3–6 June 1996; Volume 4, pp. 2160–2165.
5. Adebisi, A.A.; Adewumi, A.O.; Ayo, C.K. Comparison of ARIMA and artificial neural networks models for stock price prediction. *J. Appl. Math.* **2014**, *2014*, 614342. [\[CrossRef\]](#)
6. Oh, K.J.; Kim, K.J. Piecewise nonlinear model for financial time series forecasting with artificial neural networks. *Intell. Data Anal.* **2002**, *6*, 175–185. [\[CrossRef\]](#)

7. Yoon, Y.; Swales, G. Predicting stock price performance: A neural network approach. In Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences, Kauai, HI, USA, 8–11 January 1991; Volume 4, pp. 156–162.
8. Kohara, K.; Ishikawa, T.; Fukuhara, Y.; Nakamura, Y. Stock price prediction using prior knowledge and neural networks. *Intell. Syst. Account. Financ. Manag.* **1997**, *6*, 11–22. [[CrossRef](#)]
9. Tsai, C.F.; Wang, S.P. Stock price forecasting by hybrid machine learning techniques. In Proceedings of the International Multiconference of Engineers and Computer Scientists (IMECS 2009), Hong Kong, China, 18–20 March 2009; Volume 1, p. 755.
10. Hadavandi, E.; Shavandi, H.; Ghanbari, A. Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowl. Based Syst.* **2010**, *23*, 800–808. [[CrossRef](#)]
11. Chen, T.L.; Cheng, C.H.; Teoh, H.J. Fuzzy time-series based on Fibonacci sequence for stock price forecasting. *Phys. A Stat. Mech. Appl.* **2007**, *380*, 377–390. [[CrossRef](#)]
12. Cheng, C.H.; Chen, T.L.; Wei, L.Y. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Inf. Sci.* **2010**, *180*, 1610–1629. [[CrossRef](#)]
13. Brailsford, T.J. The empirical relationship between trading volume, returns and volatility. *Account. Financ.* **1996**, *36*, 89–111. [[CrossRef](#)]
14. Gwilym, O.A.; McMillan, D.; Speight, A. The intraday relationship between volume and volatility in LIFFE futures markets. *Appl. Financ. Econ.* **1999**, *9*, 593–604. [[CrossRef](#)]
15. Brooks, C. Predicting stock index volatility: Can market volume help? *J. Forecast.* **1998**, *17*, 59–80. [[CrossRef](#)]
16. Le, V.; Zurbruegg, R. The role of trading volume in volatility forecasting. *J. Int. Financ. Mark. Inst. Money* **2010**, *20*, 533–555. [[CrossRef](#)]
17. Tsang, W.W.H.; Chong, T.T.L. Profitability of the on-balance volume indicator. *Econ. Bull.* **2009**, *29*, 2424–2431.
18. Nedunchezian, V.R. Analysis of on Balance Volume in MCX Energy. *J. Appl. Manag. Res.* **2014**, *3*, 15–23.
19. Szűcs, B.Á. Forecasting intraday volume: Comparison of two early models. *Financ. Res. Lett.* **2017**, *21*, 249–258. [[CrossRef](#)]
20. Tsinaslanidis, P.E.; Kugiumtzis, D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Syst. Appl.* **2014**, *41*, 6848–6860. [[CrossRef](#)]
21. Nakagawa, K.; Imamura, M.; Yoshida, K. Stock Price Prediction with Fluctuation Patterns Using Indexing Dynamic Time Warping and k^* -Nearest Neighbors. In *JSAI International Symposium on Artificial Intelligence*; Springer: Cham, Switzerland, 2017; pp. 97–111.
22. Tsinaslanidis, P.E. Subsequence dynamic time warping for charting: Bullish and bearish class predictions for NYSE stocks. *Expert Syst. Appl.* **2018**, *94*, 193–204. [[CrossRef](#)]
23. Kim, S.H.; Lee, H.S.; Ko, H.J.; Jeong, S.H.; Byun, H.W.; Oh, K.J. Pattern matching trading system based on the dynamic time warping algorithm. *Sustainability* **2018**, *10*, 4641. [[CrossRef](#)]
24. Allen, F.; Karjalainen, R. Using genetic algorithms to find technical trading rules. *J. Financ. Econ.* **1999**, *51*, 245–271. [[CrossRef](#)]
25. Kim, K.J.; Han, I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Syst. Appl.* **2000**, *19*, 125–132. [[CrossRef](#)]
26. Ahn, J.J.; Kim, Y.M.; Yoo, K.; Park, J.; Oh, K.J. Using GA-Ridge regression to select hydro-geological parameters influencing groundwater pollution vulnerability. *Environ. Monit. Assess.* **2012**, *184*, 6637–6645. [[CrossRef](#)] [[PubMed](#)]
27. Kang, J.; Lee, H.J.; Jeong, S.H.; Lee, H.S.; Oh, K.J. Developing a Forecasting Model for Real Estate Auction Prices Using Artificial Intelligence. *Sustainability* **2020**, *12*, 2899. [[CrossRef](#)]
28. Song, H.; Han, S.K.; Jeong, S.H.; Lee, H.S.; Oh, K.J. Using Genetic Algorithms to Develop a Dynamic Guaranteed Option Hedge System. *Sustainability* **2019**, *11*, 4100. [[CrossRef](#)]
29. Kim, J.; Shin, S.; Lee, H.S.; Oh, K.J. A Machine Learning Portfolio Allocation System for IPOs in Korean Markets Using GA-Rough Set Theory. *Sustainability* **2019**, *11*, 6803. [[CrossRef](#)]
30. Choi, J.H.; Larsen, K.; Seppi, D.J. Equilibrium Effects of Intraday Order-Splitting Benchmarks. *Math. Finan. Econ.* **2020**. [[CrossRef](#)]
31. Konishi, H. Optimal slice of a VWAP trade. *J. Financ. Mark.* **2002**, *5*, 197–221. [[CrossRef](#)]
32. Kakade, S.M.; Kearns, M.; Mansour, Y.; Ortiz, L.E. Competitive algorithms for VWAP and limit order trading. In Proceedings of the 5th ACM Conference on Electronic Commerce, New York, NY, USA, 17–20 May 2004; Association for Computing Machinery: New York, NY, USA, 2004; pp. 189–198.
33. McCulloch, J.; Kazakov, V. *Optimal VWAP Trading Strategy and Relative Volume*; University of Technology: Sydney, Australia, 2007.
34. Białkowski, J.; Darolles, S.; Le Fol, G. Improving VWAP strategies: A dynamic volume approach. *J. Bank. Financ.* **2008**, *3*, 1709–1722. [[CrossRef](#)]
35. Fuh, C.D.; Teng, H.W.; Wang, R.H. On-line VWAP trading strategies. *Seq. Anal.* **2010**, *29*, 292–310. [[CrossRef](#)]
36. Cartea, Á.; Jaimungal, S. A closed-form execution strategy to target volume weighted average price. *SIAM J. Financ. Math.* **2016**, *7*, 760–785. [[CrossRef](#)]
37. McCulloch, J.; Kazakov, V. Mean Variance Optimal VWAP Trading. Available online: <https://ssrn.com/abstract=1803858> (accessed on 4 October 2020).
38. Insights, F. *Marching up the Learning Curve: The Second Buy Side Algorithmic Trading Survey (Financial Insights Special Report)*; Bank of America: Charlotte, NC, USA, 2006.
39. Keogh, E.J.; Pazzani, M.J. Derivative dynamic time warping. In Proceedings of the 2001 SIAM International Conference on Data Mining, Chicago, IL, USA, 5–7 April 2001; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2001; pp. 1–11.

40. Berndt, D.J.; Clifford, J. Using dynamic time warping to find patterns in time series. In *KDD workshop* **1994**, *10*, 359–370.
41. Müller, M. Dynamic time warping. In *Information Retrieval for Music and Motion*; Springer: Berlin, Germany, 2007; pp. 69–84.
42. Coelho, M.S. Patterns in Financial Markets: Dynamic Time Warping. Ph.D. Thesis, NSBE-UNL, Carcavelos, Portugal, 2012.
43. Keogh, E.; Ratanamahatana, C.A. Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **2005**, *7*, 358–386. [[CrossRef](#)]
44. Tsoukalas, L.H.; Uhrig, R.E. *Fuzzy and Neural Approaches in Engineering*; Wiley: Hoboken, NJ, USA, 1997.
45. Ahn, J.J.; Byun, H.W.; Oh, K.J.; Kim, T.Y. Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Syst. Appl.* **2012**, *39*, 8369–8379. [[CrossRef](#)]
46. Cheong, D.; Kim, Y.M.; Byun, H.W.; Oh, K.J.; Kim, T.Y. Using genetic algorithm to support clustering-based portfolio optimization by investor information. *Appl. Soft Comput.* **2017**, *61*, 593–602. [[CrossRef](#)]