Predicting Index Returns from the Market Structure Disagreement: Evidence from China

Zhipeng Ge, Wenpeng Wang, Dafeng Chen

***Abstract*—The factors with information are key to predict stock returns. Previous studies on disagreement mainly focus on the report of the analysts and the sentiment or belief of investors, as well as the trading volume or turnover, ignore the structure among stocks. In this paper, we come up with a new concept of the market structure disagreement and measure it based on the K-means clustering algorithm and Gini impurity. The experiments for the CSI100 stock market show that the market structure disagreement could improve the predicting direction accuracy of machine learning algorithms nearly 1.5%. Specifically, trading volumes and net capital inflows affect the market structure disagreement, which increases with the log difference of trading volumes and decreases with the growth rate of net capital inflows. This paper proposes a new information factor, structure disagreement, which is significantly helpful for investors with market timing, especially for the investors using machine learning.**

***Index Terms*—structure disagreement, machine learning; index forecasting.**

I. Introduction

IT is a challenging problem to understand stock market fluctuation. In the modern finance, efficient market hypothesis proposed by Malkiel and Fama [1] occupies an important position. The efficient market hypothesis holds that in the stock market with sound law, good function, high transparency and sufficient competition, all valuable information has been timely, accurately, and fully reflected in the stock price, including the current and future value of the enterprise. Unless there is market manipulation, investors can not obtain the excess higher than the average market level by analyzing the past prices. Under the framework of this theory, two kinds of investment strategies, active and passive strategies, are derived. When the market is fully efficient, passive investment strategy is adopted [2], while active investment strategies based on fundamental and technical aspects are choose. In the active and passive investment strategies, active investment strategy is an important way to verify the inefficiency of the stock market [3], [4].

The predictability of stock markets is also the direct evidence against the efficient market hypothesis. Meanwhile, more and more studies show that stock prices could be predicted by the factors with information [5], [6], [7]. The main prediction models include linear regression models based on factors [8], [9], autoregression models based on time series [10], [11], and nonlinear classification or regression models based on machine learning [12], [13]. All of the models have their own characteristics. The linear regression model focuses on economic and financial explanations, autoregression and classification models have higher predicting power than linear regression models and have drawn more attention [10], [14], [15], [16].

Factors with information are critical for prediction models. There are four kinds of factors that we usually used in study and practice: (1) macroeconomic factors, including exchange rate, interest rate, money supply, inflation, commodity prices, industrial production, etc. [17], [18]; (2) fundamental factors, including cash flow, scale, profitability, growth factors, value factors, etc. [19], [20]; (3) technical factors, including trading volume, turnover rate, momentum, moving average, similar moving average, homeopathic factor, relative strength factor, etc. [21], [22]; (4) other factors, including international stock market factors, sentiment factors, and other unclassified factors.

Among these factors, the disagreement has attracted more and more researchers' attention. Disagreement means that people have different thoughts and opinions for the same thing or event. Previous studies have shown that: (1) due to market disagreement, there are trading behaviors and price fluctuation in the stock market; besides, market disagreement also has a critical impact on returns, trading volumes, and risks. Hong and Stein [23] studied the disagreement, heterogeneity of investor beliefs, and found that the disagreement would cause changes in trading volumes and stock prices by analyzing the relations among the disagreement, stock prices, and trading volumes. In terms of market returns, Baker et al. [24] investigated the disagreement, differences of opinion among investors, and found that the disagreement changed the equilibrium price and increased the randomness of stock returns. In terms of trading volumes, Carlin et al. [25] believed that the disagreement, differences of opinion among mortgage dealers, would cause an increase in market trading volumes. In terms of market risks, Hong and Stein [26] found that the disagreement between bullish and bearish investors led to a large number of transactions and market stampedes under short-selling constraints when the market was down, but the disagreement had just a temporary impact on stocks [27]. (2) The disagreement could cause the mispricing of stocks. Sadka and Scherbina [28] studied the disagreement of analysts and found that the disagreement can lead to long-term mispricing of stocks. (3) The disagreement can help to predict returns effectively. As for stocks, Diether et al. [29] found that the greater disagreement, the difference between analysts' expected returns, would lead to lower returns. Cen et al. [30] regarded the standard deviation of analysts' expected stock returns as a measurement of the

Manuscript received December 24, 2019; revised July 28, 2020. This work is supported by the Major Program of National Fund of Philosophy and Social Science of China under Grant 18ZDA088, the Humanities and Social Science Fund of Ministry of Education of China under Grant 17YJCZH022.

Zhipeng Ge is with the Business School, Henan University, Kaifeng 475004, China (e-mail: lzgezhipeng@163.com).

Wenpeng Wang is with the School of Public Economics and Administration, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: wangwenpeng0214@126.com).

Dafeng Chen is with the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: 209423@nau.edu.cn).

disagreement, and found that change in the disagreement can predict future returns when the mean of analysts' expected stock returns were negative. Regarding stock markets, Park et al. [31] used the standard deviation of the analysts' expected index returns as a measurement of the disagreement and found that the disagreement can also predict index returns. Contrary to Park et al. [31], Yu [32] used the weighted standard deviation of the analyst expected returns on stocks as a measure of the disagreement on the indices, and found that this disagreement can lead to a decrease of index returns and can be replaced by the turnover rate. Referring to the previous studies about stocks or aggregate stocks, the disagreement is mainly calculated based on the trading volumes, turnover rate [23], [32], analysts' opinions [29], [26], [28], investors' sentiment [33], [34], [35], investors' belief [36], [37], and so on [27].

However, the disagreement research for aggregate stocks is lack of the consideration of market structure, which refers to relations among stocks. As we know, the stock market structure often shows stock clusters, which impact the strategies about how to choose portfolios and investment inevitably [38], [39]. Take a simple example. If there are six stocks in the market, there would be different kinds of stock clusters. In Fig. 1 (a), these stocks have similar including trading volume and turnover rate, etc. over a period and have close relationships with each other, so there is only one stock cluster covering total six stocks. That is to say, all of the stocks are consistent in the market structure, and the disagreement among clusters is low. On the contrary, in Fig. 1(d), each stock belongs to own cluster, and the stocks perform completely inconsistent with each other, so the disagreement among clusters is high. Similarly, we can know that the disagreement like Fig. 1(d) is the highest, followed by Fig. 1(c), Fig. 1(b), then Fig. 1(a). In this paper, the inconsistency among stock clusters is defined as the structure disagreement. Our concern is how to measure it and what about its effect on the prediction of index returns. Besides, what kinds of factors affect the structure disagreement?

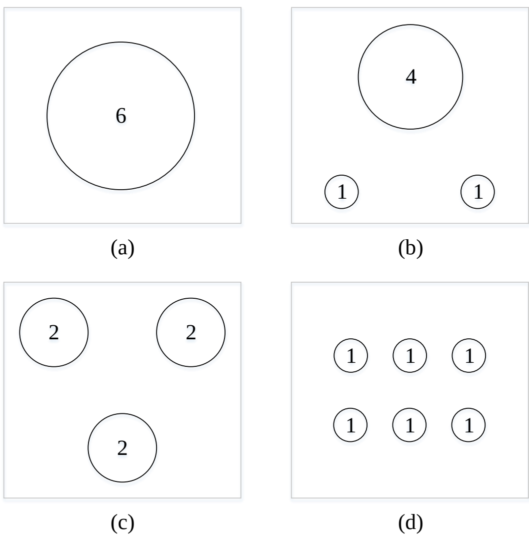

Fig. 1. Example of structure disagreement

In this paper, at first, we introduce a new concept of structure disagreement and measure it depend on the Gini impurity of clusters based on the K-means algorithm in aggregate stock markets. Then, we analyze the prediction of index returns by machine learning algorithms from the structure disagreement. Finally, we explore what kinds of factors that would affect the structure disagreement. The rest of this paper is organized as follows. In section 2, the research methods are introduced. In section 3, we introduce the concept of structural disagreement and design its measurement method. In section 4, we analyze the effect of structure disagreement on the prediction of machine learning and check its robustness. In section 5, we explore the economic interpretation of structure disagreement. In the last section, conclusions and discussions are summarized.

II. Methodology

In this paper, the clustering and regression algorithms are related. The clustering algorithm is used to mine stock clusters and calculate the disagreement, and the regression algorithm is used to predict index returns.

A. Clustering Algorithm

For clustering algorithms, we use the Kmeans algorithm, which is the most classical clustering algorithm. It is proposed by MacQueen [40] and has a good performance in many areas [41], [42], [43]. The process of this algorithm is as follows. First, initialize the number of clusters, k. Second, randomly select k objects as the centroid of clusters. Third, calculate the distance to the centroid for each node. Fourth, set the cluster for each node by its nearest centroid. Fifth, recalculate the centroids of each cluster by the mean value of its members. Sixth, re-perform the process from Third to Fifth until either cluster members do not change or the algorithm reaches its maximum number of iterations.

B. Regression Algorithm

There are many regression and classification algorithms in machine learning[44], [45], [46]. For regression algorithms, Linear Regression (LR), AdaBoost (AB), GradientBoosting (GB), XGBoost (XGB), and RandomForest (RF) are used in this paper. Linear regression is one of the most commonly used method in the financial investment. AdaBoost, GradientBoosting, RandomForest, and XGBoost algorithms are ensemble methods, which perform well and have a good robustness in many fields in recent years.

The Linear Regression (LR) model is shown as $\mathbf{y} = \mathbf{w}\mathbf{x} + b$, in which $\mathbf{w}$ and b are the coefficient vector and intercept. $\mathbf{x}$ and $\mathbf{y}$ are explanatory and response variables. Given the $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{w}$ and b can be estimated by the ordinary least squares (OLS). For detailed introduction of LR, refer to [47].

The Ada Boost (AB) is to fit a sequence of weak decision trees on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote to produce the final prediction. Initially, the weight of each iteration is set to equal. Then, wrong samples are punished and the weight of each iteration is modified according to this model error. Finally, these weak decision trees are combined. For detailed introduction of AdaBoost Regression, refer to [48].

The Gradient Boosting (GB) is proposed by Friedman [49] and contains a series of weak learners, Gradient Boosting

Trees. Like the AdaBoost, it's the first to train a series of Gradient Boosting Trees by boosting and then combine these weak learners. Because the weight is adjusted by wrong samples, the AB and GB can be significantly affected by noise or outliers.

The Random Forest (RF) is a classification technique developed by Breiman [50] and also contains some learners, Decision Trees. However, unlike the AdaBoost and GradientBoosting, these learners are randomly selected variables and trained parallelly. Finally, the prediction is the sum of each learners. So, the combined prediction trees can improve the accuracy and stability of the model performances.

The XGBoost (XGB) is proposed by Chen and Guestrin [51] and composed by a set of classification and regression trees (CART). Like Random Forest, the final prediction is combined by these CART, which are randomly selected variables. Because all meta-trees are singly trained, the RF and XGB are not sensitive to noise or outliers.

III. Measuring structure disagreement

The market structure disagreement (SD) is defined as the inconsistency among stock clusters within a period, given the stock characteristics. To measure this structure disagreement, firstly, we have to find out available stock clusters in the market, then find the way to measure this incondistency among stock clusters.

Following this thought, firstly, we use the Kmeans clustering algorithm, which is the most classical and has a good performance in many areas, to recognize stock clusters. Considering the technical factors are used in the article and stocks are often divided into odd categories, including up, down and sideways [52], or abnormal up, abnormal down, up, down and sideways, we set the number of clusters as 3, 5, and 7, respectively. Meanwhile, we set the random seeds from 0 to 29 with interval 1 to reduce the impact of randomness, which means there are 30 experiments for the Kmeans algorithm.

Then, we use Gini impurity which is often used in classification problems to measure the inconsistency among clusters [53], [54]. The Gini impurity is a measurement of how often an element randomly chosen from a set is incorrectly labeled if it is randomly labelled according to the distribution of labels in this set. The Gini impurity is calculated as

$$G = \sum_{i=1}^{i=C} f_i * (1 - f_i) \tag{1}$$

Where the G and C represent the Gini impurity and the number of clusters, and the f_i is the probability of a stock belongs to the ith cluster. From the Eq. 1, we can know that the structure disagreement is 0,0.5,0.67, and 0.83 in Fig. 1 (a), (b), (c), and (d) respectively. The structure disagreement in Fig. 1(d) is the largest, followed by Fig. 1(c), Fig. 1(b), then Fig. 1(a).

A. The measure of structure disagreement

In this part, we take a specific example to introduce how to get the market structure disagreement for CSI 100 stock marekts. The component stock data is downloaded on Feb. 28, 2019, and all data comes from the CHOICE database (http://www.eastmoney.com). In terms of time periods, the data ranges from Jul. 13, 2015, to Jul. 19, 2015, for one week. As for stock characteristics, we use the log difference of stock daily trading volumes within a week, which means there are 5 clustering characteristics when we analyze stock clusters, and set the number of clusters and the random seed as 5 and 0, respectively.

The clusters of the stocks are showing in Table I, for CSI 100 stock markets from Jul. 13, 2015, to Jul. 19, 2015, there are 5 clusters, and the SD is 0.366. In this period, most of the stocks have similar characteristics and trends in the log difference of trading volumes as in Fig. 1(b), and the structure disagreement is low.

TABLE I
THE STOCK CLUSTERS FOR CSI100 STOCK MARKETS FROM JUL. 13, 2015, TO JUL. 19, 2015, WHEN THE RANDOM SEED IS 0.

Clusters	Stocks
0	000063.
1	000725, 600703.
2	000651, 002024, 002352, 002450, 002594, 600023, 600115, 600518, 600606, 601018, 601633, 601727, 601933, 601985, 601989, 603993.
3	000069.
4	000001, 000002, 000166, 000333, 000538, 000568, 000776, 000858, 000895, 002142, 002252, 002304, 002736, 300059, 600000, 600009, 600010, 600011, 600015, 600016, 600018, 600019, 600028, 600030, 600036, 600048, 600050, 600104, 600276, 600309, 600340, 600519, 600585, 600690, 600837, 600887, 600958, 600999, 601006, 601009, 601088, 601166, 601169, 601186, 601211, 601225, 601238, 601288, 601318, 601328, 601336, 601360, 601390, 601398, 601601, 601618, 601628, 601668, 601669, 601688, 601766, 601800, 601818, 601857, 601888, 601899, 601988, 601998, 603288.
SD	16/89*(1-16/89) + 69/89*(1-69/89) + 2/89*(1-2/89) + 1/89*(1-1/89) + 1/89*(1-1/89) =0.366

Then, to reduce the impact of random seed on the Kmeans clustering algorithm, we implement 30 independent experiments on the Kmeans algorithm. The detailed results are shown in Table II. We use the average value of 30 independent experiments as the structure disagreement, which means the structure disagreement is 0.374 for CSI 100 stock markets from Jul. 13, 2015, to Jul. 19, 2015.

B. The statistics of structure disagreement

Based on the above calculation, we also use the log difference of daily trading volumes to recognize the stock clusters. Then, we calculate the structure disagreement of CSI 100 stock markets from Jan. 15, 2007, to Jul. 1, 2018, 559 weeks. In this period, the CSI experienced abnormal rise and fall around 2008 and 2015, the slow decline from 2009 to 2013, the sideways in 2014, and the slow rise from 2016 to 2018, which were collected on Feb. 28, 2019, and included most of the patterns to test. Fig. 2 and 3 represently show the time series of log return of CSI 100 and the structure disagreement with the number of clusters 3, 5, and 7 from Jan. 15, 2007, to Jul. 1, 2018.

In Fig. 3, it can be seen that the SD of CSI 100 stock markets is particularly abnormal from Apr. to Aug. in 2015. Similarly to the market structure in Fig. 1 (b), most of the

TABLE II
THE MARKET STRUCTURE DISAGREEMENT FOR CSI100 STOCK MARKETS FROM JUL. 13, 2015, TO JUL. 19, 2015.

	7/13/2015 to 7/19/2015									
Seeds	0	1	2	3	4	5	6	7	8	9
SD	0.366	0.379	0.366	0.379	0.366	0.379	0.392	0.366	0.379	0.379
Seeds	10	11	12	13	14	15	16	17	18	19
SD	0.379	0.366	0.379	0.366	0.366	0.379	0.366	0.379	0.379	0.379
Seeds	20	21	22	23	24	25	26	27	28	29
SD	0.366	0.379	0.379	0.366	0.379	0.366	0.379	0.379	0.379	0.366
avg	0.374									

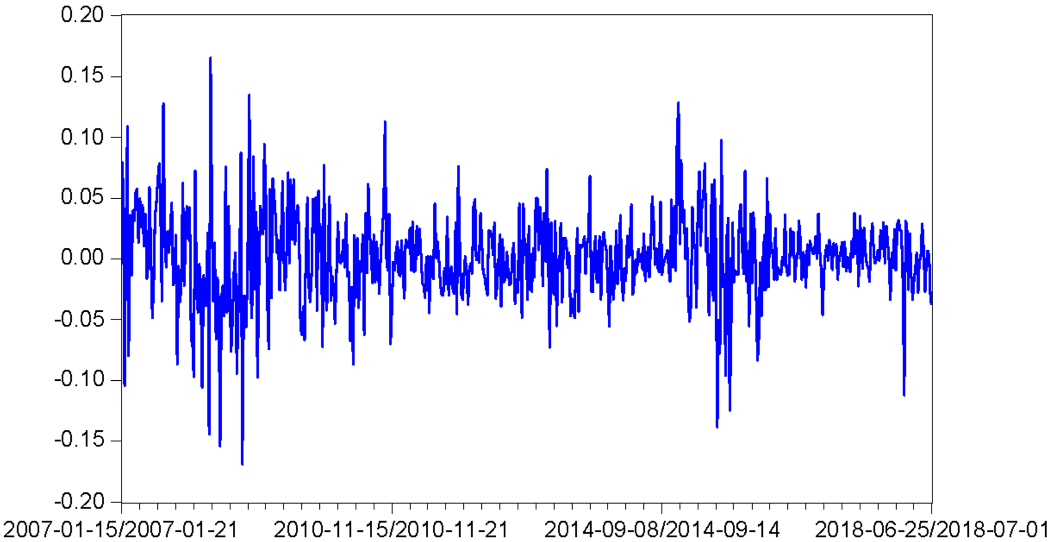

Fig. 2. The weekly log return of CSI 100 from Jan. 15, 2007, to Jul. 1, 2018.

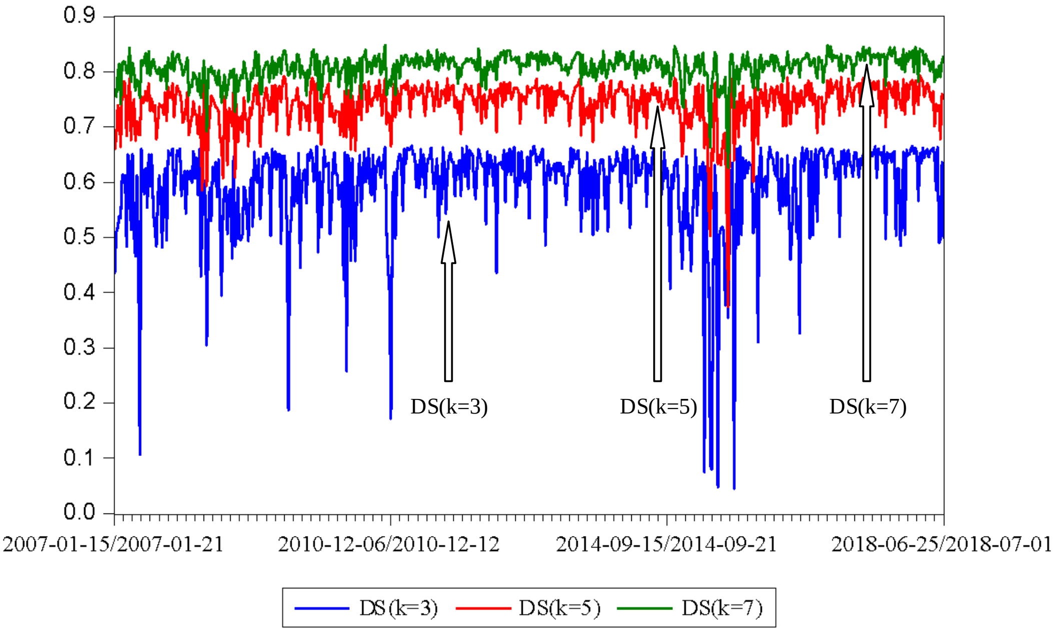

Fig. 3. The weekly SD of CSI 100 stock markets from Jan. 15, 2007, to Jul. 1, 2018.

stocks have similar characteristics and trends and belong to one cluster from Apr. to Aug. in 2015, where the CSI 100 indices are sideways and falling.

Table III shows the statistics of the market structure disagreement for CSI 100 stock markets from Jan. 15, 2007 to Jul. 1, 2018. The mean of the structure disagreement increases with the number of clusters, which can also be seen in Fig. 3. However, the standard deviation of structure disagreement decreases with the number of clusters, which means that the information of the market structure disagreement decreases with the number of clusters. Especially, the smallest correlation coefficient among SD with different cluster numbers reaches 0.610, the value is so high that we would not study other cluster numbers for recognizing clusters further.

IV. Predicting index returns from stock market structure

A lot of research shows that machine learning algorithms have an excellent performance in the prediction of returns [12], [6]. While using machine learning to predict index returns, there are four problems we have to pay more attention. (1) After standardizing or normalizing the data, the training and test set is directly divided, so there is information on the test set in the training set. (2) The training and the test set are often randomly divided, which can also lead to the use of future information in the training set. (3) There are periodic problems in the training set. The frequency of training is not the higher, the better. Factors may contain different periodic characteristics, which can lead to the evaluation error in the valuation, so the research often ignores the factor periodicity. (4) There is not any interpretability for selected factors when using machine learning to predict index returns. The interpretability of factors, which determines the rationality of factors and their economic implications, is a particular concern for financial investors. Based on above four problems, we try some new ways to predict the index returns. To avoid the future information, we divide the training and test set by the calendar, standardize the test set by the standardizer of the training set, analyze the prediction of index returns from market structure disagreement in different prediciton periods, and interpret this disagreement.

A. Data description

This study uses the weekly data of CSI 100 indices and constituent stocks, which also come from the CHOICE database. The sample data is from Jan. 15, 2007, to Jul. 1, 2018, 559 samples. The data from Jan. 15, 2007, to Dec. 26, 2016, are used as the training set, and the data from Dec. 26, 2016, to Jul. 1, 2018, are used as the test set. Besides, to avoid bringing effect caused by future information in the training set, we have used the mean and standard deviation of training set to standardize the training set and test set.

In terms of dependent variables, we adopt log returns of the index after the next one week (R(1)), one month (R(4)), two months (R(8)), and one quarter (R(13)), half-year (R(26)). In terms of explanatory variables, we select 13 variables, including 6 external variables and 7 internal variables. External variables are exchange rate (USDCNY), France indices (CAC40), Germany indices (DAX), S&P indices (SP500), Hang Seng indices (HSI), and Nikkei 225 indices (N225) and they are all from CHOICE database. Internal variables are net capital inflows (NCI), commodity channel index (CCI)), moving average convergence divergence (MACD), momentum (MOM), relative strength index (RSI), simple moving average (SMA), and William index (WillR) and they are all calculated by Ta-lib (http://www.ta-lib.org). The special or newly added variables include volumes (V) and the market structure disagreement (SD), in which the trading volume is often used as a proxy of the disagreement and can be used to compare with SD [23]. In order to ensure the data stability, we use the growth rate of net capital inflows (NCI*) to replace the net capital inflows and use the log difference of variables to replace other variables, including the log difference of volumes, simple moving average, exchange rate, France indices, Germany indices, S&P indices, Hang Seng indices, and Nikkei 335 indices (V*, SMA*, USDCNY*, CAC40*, DAX*, SP500*, HSI*, and N225*). The statistics of all variables are shown in Table IV. They are all stable time series according to the ADF test.

B. Prediction models

In this paper, Linear Regression (LR), Ada Boost (AB), Gradient Boosting (GB), XGBoost (XGB), and Random Forest (RF) are used to predict the stock index. It should be noted that, in addition to linear regression, other four machine learning methods are susceptible to the random seed. So we implement 30 times experiments for the four algorithms and use the mean of 30 experiments as the prediction performance to reduce the effect of the random seed.

To check the effect of the proposed structure disagreement on the prediction of index returns based on machine learning, two benchmark models are constructed. Model1 only uses explanatory variables, including 13 variables, and Model2 uses explanatory variables and the trading volumes (V*), including 14 variables. Besides, we construct two additional models. Model3 uses explanatory variables and the structure disagreement (SD), including 14 variables, and Model4 uses explanatory variables, volumes, and the structure disagreement, including 15 variables. The input characteristics or variables of different models are shown in Table V.

C. Evaluation criteria

The direction of index movement is important to finance investors, so we mainly use the direction accuracy (DA) as evaluation criteria to assess the judgment. The calculation of DA is as follows

$$DA = \frac{1}{N}\sum_{i=1}^{N} \max(0, sign(y_i * \hat{y}_i)), \tag{2}$$

where, N is the number of objects in the test set, y_i and $\hat{y}_i$ represent the real and predicting log returns for object i. $sign(y_i * \hat{y}_i)$ is the sign function. If $y_i * \hat{y}_i$ is greater than 0, it's 1, otherwise it's 0.

TABLE III
THE STATISTICS OF MARKET STRUCTURE DISAGREEMENT FOR CSI 100 STOCK MARKETS FROM 1/15/2007 TO 7/1/2018.

SD	Mean	Std	Min	Quantile 25%	Quantile 50%	Quantile 75%	Max	SD(k=3)	SD(k=5)	SD(k=7)
SD(k=3)	0.593	0.088	0.045	0.569	0.623	0.648	0.666	1.000	0.655	0.610
SD(k=5)	0.742	0.041	0.374	0.726	0.750	0.770	0.794	0.655	1.000	0.861
SD(k=7)	0.809	0.025	0.627	0.798	0.813	0.826	0.849	0.610	0.861	1.000

TABLE IV
THE STATISTICS OF VARIABLES FOR CSI 100 STOCK MARKETS FROM 1/15/2007 TO 7/1/2018.

Variable	Mean	Std	Min	Quantile 25%	Quantile 50%	Quantile 75%	Max	ADF(p)
R(1)	0.001	0.038	-0.154	-0.021	0.000	0.022	0.165	0.000
R(4)	0.002	0.081	-0.270	-0.042	0.002	0.043	0.322	0.000
R(8)	0.002	0.122	-0.357	-0.064	0.002	0.063	0.426	0.000
R(13)	0.002	0.163	-0.503	-0.090	0.004	0.086	0.474	0.001
R(26)	0.000	0.259	-0.855	-0.143	-0.008	0.121	0.733	0.047
USDCNY*	0.000	0.003	-0.012	-0.001	0.000	0.001	0.029	0.000
CAC40*	0.000	0.031	-0.251	-0.016	0.003	0.018	0.124	0.000
DAX*	0.001	0.032	-0.244	-0.015	0.004	0.018	0.149	0.000
SP500*	0.001	0.025	-0.201	-0.009	0.003	0.014	0.114	0.000
HSI*	0.001	0.031	-0.178	-0.018	0.003	0.020	0.117	0.000
N225*	0.001	0.032	-0.279	-0.015	0.003	0.019	0.115	0.000
NCI*	-0.525	51.620	-954.250	-0.776	-0.211	0.742	708.947	0.000
CCI	7.077	95.021	-166.667	-80.703	14.104	94.890	166.667	0.000
MACD	11.612	154.874	-495.674	-58.357	4.770	72.289	577.057	0.000
MOM	15.639	318.340	-1046.754	-128.728	20.247	161.739	1311.986	0.000
RSI	52.533	21.844	11.220	36.356	51.776	68.598	99.883	0.000
SMA*	0.001	0.019	-0.060	-0.010	0.002	0.010	0.068	0.000
WillR	46.716	31.688	0.000	17.267	45.392	77.824	100.000	0.000
V*	0.019	0.394	-1.499	-0.193	-0.001	0.184	2.301	0.000
SD(k=3)	0.593	0.088	0.045	0.569	0.623	0.648	0.666	0.000
SD(k=5)	0.742	0.041	0.374	0.726	0.750	0.770	0.794	0.000
SD(k=7)	0.809	0.025	0.627	0.798	0.813	0.826	0.849	0.000

TABLE V
INPUTING VARIABLES OF DIFFERENT MODELS FOR MACHINE LEARNING.

Model	Variable
Model1	USDCNY*, CAC40*, DAX*, SP500*, HSI*, N225*, NCI*, CCI, MACD, MOM, RSI, SMA*, WillR
Model2	USDCNY*, CAC40*, DAX*, SP500*, HSI*, N225*, NCI*, CCI, MACD, MOM, RSI, SMA*, WillR, **V***
Model3	USDCNY*, CAC40*, DAX*, SP500*, HSI*, N225*, NCI*, CCI, MACD, MOM, RSI, SMA*, WillR, **SD**
Model4	USDCNY*, CAC40*, DAX*, SP500*, HSI*, N225*, NCI*, CCI, MACD, MOM, RSI, SMA*, WillR, **V***, **SD**

D. Results

1) Using the log difference of daily trading volumes to calculate the SD: As for the measurement of structure disagreement above, we first use the log difference of daily trading volumes to measure weekly structure disagreement. In terms of the machine learning, we use the default hyper-parameters in the sklearn package (https://scikit-learn.org). The $Z-$score standardization is trained by the training set and then used in test set to reduce the influence of different magnitude on the gradient because Ada Boost, Gradient Boosting, and XGBoost involve the conduction of gradients. We use SD (k=5) as an example to illustrate the experimental results in this section. The training data are from Jan. 15, 2007, to Dec. 26, 2016, which is used to train different machine learning models. The test data are from Dec. 26, 2016, to Jul. 1, 2018, to test the performance of different models.

TableVI and VII report the detailed accuracy of each model and the p-values of one-sided t-test between different models. From these tables, we find that model2 does not show a significant advantage in accuracy compared with model1. The average accuracy of model1 is 0.582, while that of model2 is 0.581, which indicates the trading volume does not increase the predicted direction accuracy. However, the model3, whose average accuracy is 0.597, has an improvement in accuracy compared with the model1, which represents the structure disagreement can improve the accuracy of machine learning models. Besides, the model4 whose average accuracy is 0.593 does not show much advantage compared with the model3, which means that involving all variables directly into machine learning models does not bring an improvement in the predicted direction accuracy. Factors have different characteristics or properties, so they can not be simply combined. For example, the trading volume can improve the predicted direction accuracy of index in the next week, but the structure disagreement improves the accuracy in the next eight weeks. So, the hybrid model, model4, does not show an advantage over other single models, model2 and model3, in the predicted direction accuracy of the indices in the next one and eight weeks.

In general, (1) Compared with trading volumes, a proxy of the disagreement, the market structure disagreement im-

TABLE VI
THE DA RESULT OF DIFFERENT MODELS IN THE TEST SET.

Weeks	Model1					Avg	Model2					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.595	0.498	0.509	0.568	0.538	0.541	0.608	0.510	0.601	0.581	0.529	0.566
4	0.635	0.590	0.550	0.568	0.497	0.568	0.635	0.590	0.518	0.514	0.501	0.552
8	0.541	0.737	0.617	0.608	0.550	0.611	0.514	0.723	0.608	0.662	0.567	0.615
13	0.541	0.764	0.649	0.635	0.573	0.632	0.527	0.759	0.673	0.635	0.576	0.634
26	0.378	0.643	0.601	0.608	0.571	0.560	0.419	0.641	0.523	0.554	0.554	0.538
Avg	0.538	0.646	0.585	0.597	0.546	**0.582**	0.541	0.644	0.585	0.589	0.545	**0.581**
Weeks	Model3					Avg	Model4					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.595	0.514	0.545	0.568	0.536	0.551	0.581	0.531	0.538	0.635	0.532	0.563
4	0.622	0.577	0.570	0.541	0.511	0.564	0.635	0.591	0.550	0.473	0.506	0.551
8	0.676	0.731	0.676	0.595	0.576	0.650	0.676	0.7.1	0.662	0.595	0.553	0.643
13	0.635	0.770	0.629	0.649	0.579	0.652	0.649	0.764	0.608	0.622	0.576	0.644
26	0.614	0.641	0.527	0.595	0.558	0.567	0.486	0.643	0.574	0.568	0.558	0.566
Avg	0.608	0.647	0.589	0.589	0.552	**0.597**	0.605	0.652	0.586	0.578	0.545	**0.593**

TABLE VII
THE MODEL STATISTICS AND THE p-VALUES OF ONE-SIDED t-TEST BETWEEN DIFFERENT MODELS.

	Min	Mean	Max	Std	p-value			
					Model1	Model2	Model3	Model4
Model1	0.378	0.582	0.764	0.076	-	0.405	**0.068**	0.145
Model2	0.419	0.581	0.759	0.075	-	-	**0.051**	0.123
Model3	**0.511**	**0.597**	**0.770**	**0.064**	-	-	-	0.144
Model4	0.473	0.593	0.764	0.069	-	-	-	-

proves the predicted direction accuracy of index. (2) Investors can not merely use a basket of factors as input variables for machine learning models because factors may have different periodic properties.

2) Using the log difference of daily stock turnover to calculate the SD: Considering trading volumes and turnover rate are critical indicators investors concerned, therefore, in this experiment, for the clustering characteristics, we use the log difference of stock daily turnover rate to replace the log difference of stock daily trading volumes, to measure the market structure disagreement. We still use SD (k=5) as an example, and the predicted direction accuracy of index is shown in Table VIII.

In Table VIII, we find that the conclusion of predicted direction accuracy (DA) is the same when using the log difference of stock daily turnover rate as the clustering characteristic, compared with the log difference of stock trading volumes. The market structure disagreement increases the predicted accuracy of machine learning models. The average accuracy of model4, 0.596, is more significant than that of model3, 0.595, which indicates that combining trading volumes and the market structure disagreement can improve the predicted accuracy of machine learning models compared with only trading volumes or the market structure disagreement, but the magnitude is small. The previous conclusion that investors can not merely use a basket of factors as input variables of machine learning algorithms is still standing.

3) Using the classic F1 score to evaluate the impact of SD on the prediction: The F1 score, which combines the precision (P) and recall (R), is the most classical indicator to evaluate the classification performance of machine learning models. Precision refers to the proportion of positive predictions for all predictions. Recall means the ratio of positive predictions to the number it should have. Generally, the increase in precision often leads to a decrease in recall. The F1 score is an indicator, which comprehensively considers the precision and recall and it is calculated as

$$F1 = \frac{2PR}{P+R}. \quad (3)$$

We still use SD (k=5) as an example and test different machine learning models for the CSI 100 stock market from Dec. 26, 2016, to Jul. 1, 2018. The F1 score result is shown in Table IX.

Compared with the result of predicted direction accuracy, the conclusion of the F1 score does not change. Compared with trading volume, the market structure disagreement improves the prediction performance of index returns based on machine learning models. Also, when using machine learning algorithms to predict index returns, we cannot merely use a basket of factors as input characteristics.

4) Using another number of clusters, 3 and 7, to calculate the SD: In this part, we analyze the effect of the market structure disagreement on the prediction of machine learning algorithms. From the correlation among different structure disagreement in Table III, we know that the correlation efficient between SD(k=5) and SD(k=7) is 0.861, which is large, so we needn't analyze the number of clusters higher than 7. So we evaluate the impact of the market structure disagreement when the number of clusters is 3 and 7. We still use the log difference of stock daily trading volumes as the clustering characteristic to measure the market structure disagreement, and the predicted direction accuracy of different models is shown in Table X.

There is a similar effect of the market structure disagreement with the number of clusters of 3, 5, and 7 on the predicted direction accuracy of machine learning models. The market structure disagreement improves the prediction accuracy compared with trading volumes. At the same time, we find that the accuracy of the structure disagreement with

TABLE VIII
THE DA RESULT OF DIFFERENT MODELS WHEN THE CLSUTER CHARACTERISTIC IS REPLACED BY STOCK DAILY TURNOVER RATE.

Weeks	Model3					Avg	Model4					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.581	0.505	0.548	0.527	0.532	0.538	0.568	0.520	0.641	0.622	0.528	0.576
4	0.635	0.586	0.572	0.554	0.505	0.570	0.635	0.599	0.506	0.500	0.516	0.551
8	0.676	0.749	0.645	0.635	0.564	0.654	0.676	0.739	0.658	0.649	0.544	0.653
13	0.622	0.778	0.666	0.689	0.575	0.666	0.635	0.764	0.659	0.595	0.568	0.644
26	0.500	0.639	0.514	0.527	0.550	0.546	0.500	0.636	0.527	0.568	0.549	0.556
Avg	0.603	0.651	0.589	0.586	0.545	**0.595**	0.603	0.651	0.598	0.586	0.541	**0.596**

TABLE IX
THE F1 SCORE RESULT OF DIFFERENT MODELS IN TEST SET.

Weeks	Model1					Avg	Model2					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.590	0.495	0.508	0.568	0.538	0.540	0.603	0.509	0.593	0.582	0.530	0.563
4	0.634	0.580	0.537	0.571	0.498	0.564	0.634	0.576	0.517	0.518	0.502	0.549
8	0.548	0.733	0.621	0.614	0.553	0.614	0.519	0.721	0.612	0.663	0.570	0.617
13	0.547	0.751	0.652	0.640	0.577	0.633	0.534	0.744	0.677	0.638	0.580	0.635
26	0.369	0.604	0.599	0.607	0.568	0.549	0.412	0.595	0.523	0.554	0.552	0.527
Avg	0.538	0.633	0.584	0.600	0.547	**0.580**	0.540	0.629	0.585	0.591	0.547	**0.578**
Weeks	Model3					Avg	Model4					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.583	0.504	0.536	0.563	0.536	0.544	0.567	0.518	0.526	0.627	0.531	0.554
4	0.590	0.560	0.568	0.538	0.507	0.553	0.607	0.576	0.544	0.471	0.504	0.540
8	0.648	0.719	0.678	0.600	0.575	0.644	0.648	0.724	0.663	0.601	0.553	0.638
13	0.631	0.754	0.634	0.653	0.582	0.651	0.646	0.748	0.614	0.627	0.579	0.643
26	0.509	0.598	0.527	0.594	0.554	0.556	0.482	0.601	0.574	0.564	0.555	0.555
Avg	0.592	0.627	0.588	0.589	0.551	**0.590**	0.590	0.633	0.584	0.578	0.544	**0.586**

TABLE X
THE DA RESULT OF DIFFERENT MODELS WHEN THE NUMBER OF CLUSTER IS 3 AND 7.

	Weeks	Model3					Avg	Model4					Avg
		LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
SD(k=3)	1	0.595	0.533	0.553	0.635	0.537	0.571	0.608	0.553	0.592	0.595	0.543	0.578
	4	0.676	0.575	0.595	0.500	0.543	0.578	0.662	0.573	0.568	0.473	0.518	0.559
	8	0.676	0.731	0.677	0.622	0.550	0.651	0.662	0.736	0.603	0.676	0.554	0.646
	13	0.662	0.767	0.635	0.662	0.588	0.663	0.649	0.777	0.673	0.622	0.571	0.658
	26	0.486	0.650	0.536	0.622	0.566	0.572	0.500	0.645	0.543	0.608	0.545	0.568
	Avg	0.619	0.651	0.599	0.608	0.557	**0.607**	0.616	0.657	0.596	0.595	0.546	**0.602**
	Weeks	Model3					Avg	Model4					Avg
		LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
SD(k=7)	1	0.595	0.541	0.519	0.581	0.543	0.556	0.595	0.553	0.579	0.608	0.539	0.575
	4	0.649	0.593	0.546	0.473	0.529	0.558	0.649	0.600	0.599	0.527	0.523	0.580
	8	0.635	0.736	0.705	0.716	0.574	0.673	0.649	0.733	0.664	0.689	0.580	0.663
	13	0.635	0.768	0.670	0.622	0.584	0.656	0.635	0.769	0.624	0.568	0.586	0.636
	26	0.459	0.645	0.505	0.554	0.571	0.547	0.473	0.645	0.513	0.486	0.555	0.534
	Avg	0.595	0.657	0.589	0.589	0.560	**0.598**	0.600	0.660	0.596	0.576	0.557	**0.598**

the number of clusters 3 is the best among other structure disagreement, which may be caused by the high variance of the structure disagreement with cluster number 3 (Table III). Therefore, the result is consistent with the experiment in Section IV with other cluster numbers, 3 and 7, to analyze the impact of the market structure disagreement on the prediction of machine learning models.

E. Robustness

From the above experiments, we know that the market structure disagreement improves the predicted direction accuracy of the index, but there are also some issues we must consider. Is this advantage still valid when the measurement of structure disagreement or the market in which they are located change?

1) Using another clustering algorithm, agglomerative method, to measure the SD: In the above experiment, the most classic clustering algorithm, Kmeans method, is used to analyze the stock clusters and measure the market structure disagreement. To avoid the influence of the algorithm, we use the agglomerative clustering algorithm to analyze the stock clusters in this experiment. The agglomerative clustering algorithm is a hierarchical method and can help investors clearly and intuitively displaying the hierarchical structure among stocks. We still use the log difference of stock daily trading volumes as the clustering characteristic, like the initial experiment.

Fig. 4 shows the hierarchical clustering dendrogram for CSI 100 stocks from Jul. 13, 2015, to Jul. 19, 2015. The horizontal and vertical axes respectively represent stock clusters and distances between stocks. The stocks from left to right in each cluster are shown in Table XI. Like stock clusters in Table I, most of the stocks are similar and the structure disagreement is very low.

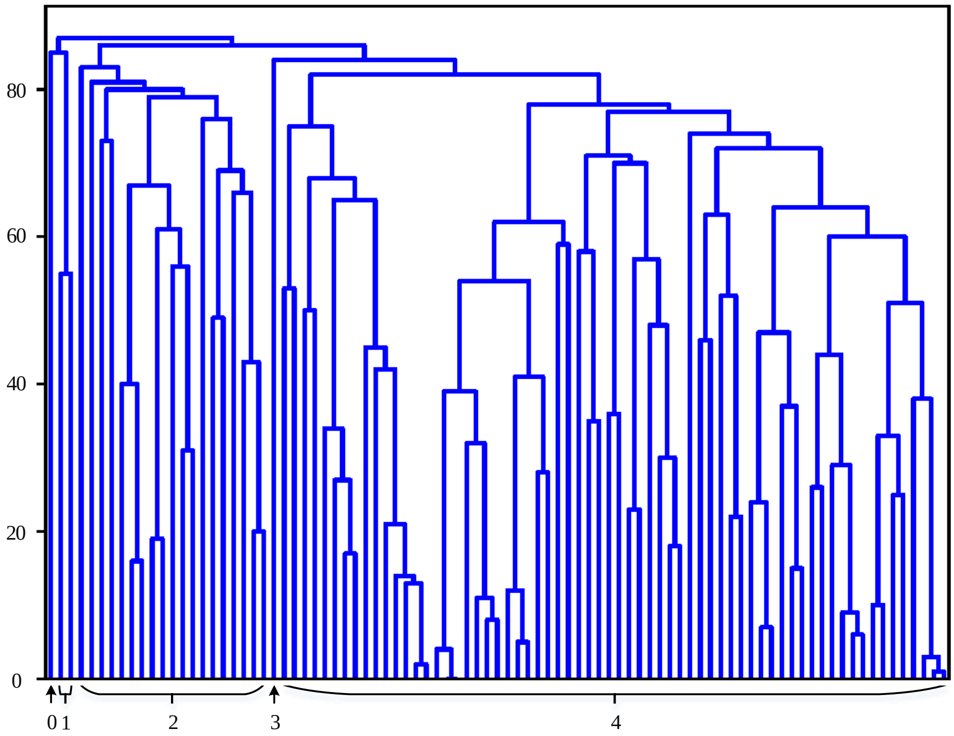

Fig. 4. The hierarchical clustering dendrogram for CSI100 stocks from Jul. 13, 2015, to Jul. 19, 2015.

TABLE XI
THE STOCK CLUSTERS BASED ON AGGLOMERATIVE CLUSTERING ALRORITHM FOR CSI 100 STOCK MARKETS FROM JUL. 13, 2015, TO JUL. 19, 2015.

Clusters	**Stocks**
0	000063.
1	000725, 600703.
2	002450, 600518, 002352, 603993, 000651, 002594, 601899, 600340, 601360, 002024, 601018, 601933, 601633, 601985, 601989, 600115, 600606, 600023, 601727.
3	000069.
4	600028, 601857, 601988, 601998, 601628, 600519, 601288, 601398, 600036, 600000, 601328, 601169, 600016, 600015, 601818, 000166, 000333, 601186, 600009, 601800, 600009, 601800, 000858, 600018, 601225, 601618, 601669, 603288, 002252, 600276, 000002, 601166, 601318, 000538, 002304, 601336, 601601, 600309, 600104, 000001, 601088, 600690, 601238, 601888, 002148, 000568, 300059, 601766, 600010, 601390, 600011, 601211, 000895, 600048, 600958, 600585, 000776, 002736, 601668, 601688, 601006, 601009, 600050, 600837, 600030, 600999.
SD(k=5)	1/89*(1-1/89)+2/89*(1-2/89)+19/89*(1-19/89) +1/89*(1-1/89)+66/89*(1-66/89)=0.404

Then, we analyze the impact of the market structure disagreement based on the agglomerative hierarchical clustering algorithm on the prediction of machine learning models. We still set the number of clusters as 5, and the predicted direction accuracy of different models shows in Table XII for CSI 100 stock market in the test set.

The result shows that the predicted direction accuracy of model3 is higher than that of model1 and model2, which also supports the conclusion we got in Section IV. Interestingly, the predicted direction accuracy based on the agglomerative clustering is better than that of Kmeans clustering algorithm. Especially, we also find using a basket of factors as input variables of machine learning algorithms is not desirable because the accuracy of model4 is worse than that of model3. In other words, we also conclude that the market structure disagreement improves the predicted direction accuracy of machine learning models, similarly with the above conclusion.

2) Using other stock market to evaluate the impact of SD on the prediction: All the above experiments are analyzed for the CSI 100 stock market. In this subsection, we discuss the effect of structure disagreement on the prediction for CSI 300 stock markets. We still use the log difference of stock daily trading volumes as the clustering characteristic, use the Kmeans clustering algorithm to recognize stock clusters, and set the number of clusters as 5. The division of training and test set is consistent with the above experiment. Table XIII and XIV represently show the predicted direction accuracy of each model and the p-values of one-sided t-test between different models in the test set for CSI 300.

As for CSI 300 stock markets, we also find that the market structure disagreement performs better than trading volumes. The accuracy of model3 is 0.565, which is larger than that of model2 and model1. Although combining trading volumes and the market structure disagreement further improve the prediction accuracy, the accuracy of model4 is slightly higher than that of the model3. So the above conclusion still stands. Besides, we find the CSI 300 market is more difficult to predict compared with the CSI 100 stock market.

TABLE XII
THE DA RESULT OF DIFFERENT MODELS BASED ON AGGLOMERATIVE CLUSTERING ALGORITHM.

Weeks	Model3					Avg	Model4					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.595	0.520	0.515	0.649	0.528	0.561	0.595	0.525	0.559	0.595	0.527	0.560
4	0.649	0.619	0.601	0.541	0.520	0.586	0.622	0.618	0.527	0.568	0.500	0.567
8	0.689	0.744	0.614	0.689	0.576	0.662	0.703	0.746	0.607	0.649	0.579	0.657
13	0.689	0.778	0.643	0.662	0.600	0.674	0.689	0.774	0.608	0.662	0.587	0.664
26	0.486	0.649	0.625	0.527	0.569	0.571	0.500	0.652	0.608	0.527	0.555	0.568
Avg	0.622	0.662	0.600	0.613	0.558	**0.611**	0.622	0.663	0.582	0.600	0.550	**0.603**

TABLE XIII
THE DA RESULT OF DIFFERENT MODELS FOR CSI 300 STOCK MARKETS.

Weeks	Model1					Avg	Model2					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.581	0.500	0.503	0.473	0.504	0.512	0.568	0.509	0.473	0.500	0.500	0.510
4	0.581	0.568	0.482	0.527	0.507	0.533	0.581	0.554	0.501	0.554	0.514	0.541
8	0.500	0.636	0.534	0.608	0.529	0.561	0.527	0.644	0.613	0.541	0.535	0.572
13	0.554	0.685	0.618	0.595	0.559	0.602	0.554	0.683	0.638	0.635	0.547	0.611
26	0.446	0.583	0.569	0.662	0.534	0.559	0.446	0.582	0.554	0.568	0.542	0.538
Avg	0.532	0.594	0.541	0.573	0.526	**0.553**	0.535	0.594	0.556	0.559	0.528	**0.554**
Weeks	Model3					Avg	Model4					Avg
	LR	AB	GB	XGB	RF		LR	AB	GB	XGB	RF	
1	0.581	0.499	0.519	0.554	0.532	0.537	0.595	0.513	0.523	0.500	0.526	0.531
4	0.595	0.608	0.550	0.554	0.504	0.562	0.581	0.621	0.568	0.514	0.499	0.556
8	0.500	0.663	0.554	0.581	0.535	0.566	0.500	0.674	0.532	0.595	0.528	0.566
13	0.595	0.685	0.651	0.581	0.573	0.617	0.595	0.682	0.676	0.635	0.583	0.634
26	0.473	0.586	0.541	0.568	0.540	0.541	0.473	0.580	0.554	0.568	0.538	0.543
Avg	0.549	0.608	0.563	0.568	0.537	**0.565**	0.549	0.614	0.570	0.562	0.535	**0.566**

TABLE XIV
THE MODEL STATISTICS AND THE p-VALUES OF ONE-SIDED t-TEST BETWEEN DIFFERENT MODELS FOR CSI 300.

	Min	Mean	Max	Std	p-value			
					Model1	Model2	Model3	Model4
Model1	0.446	0.553	0.685	0.059	-	0.440	**0.051**	**0.037**
Model2	0.446	0.554	0.683	0.055	-	-	**0.049**	**0.050**
Model3	**0.473**	0.565	**0.685**	**0.050**	-	-	-	0.384
Model4	0.473	**0.566**	0.682	0.057	-	-	-	-

In general, we also find that: (1) Compared with trading volumes, the market structure disagreement improves the predicted direction accuracy. (2) Each variable has its own property, and not more is better for machine learning models. Also, we can get some meaningful revelations. For example, the CSI 300 stock market is more difficult to predict compared with the CSI 100 stock markets.

V. THE ECONOMIC INTERPRETATION OF MARKET STRUCTURE DISAGREEMENT

Hereto, it can be found that structure disagreement improves the predicted direction accuracy of the machine learning models, but the economic implication of factors is still vital for investors. So we further demonstrate the structure disagreement from the external and internal perspectives of markets.

Unlike the above experiment, we use all data from Jan. 15, 2007, to Jul. 1, 2018, for CSI 100 stock markets to study. The dependent variable is the market structure disagreement. We use the log difference of trading volumes as the clustering characteristic and use the Kmeans algorithm to recognize stock clusters. The explanatory variable includes the exchange rate (USDCNY*), French index (CAC40*), German index (DAX*), S&P index (SP500*), Hang Seng index (HSI*), Nikkei 225 index (N225*), net capital inflows (NCI*), commodity channel index (CCI)), moving average convergence divergence (MACD), momentum (MOM), relative strength index (RSI), simple moving average (SMA*), William index (WillR), and volumes (V*) and log returns of the index (R*). The statistics of variables are shown in Table III.

Firstly, we analyze the significance of each variable by the regression model with the single variable and intercept term. The p-value result is shown in Table XV.

TABLE XV
THE p-VALUES OF REGRESSION MODELS WITH EACH SINGLE EXPLAIN VARIABLE.

Category	p-value of variables					
External	Variable	USDCNY*	CAC40*	DAX*		
	p-value	0.19	0.28	0.19		
	Variable	SP500*	HSI*	N225*		
	p-value	0.11	**0.05**	**0.02**		
Internal	Variable	NCI*	CCI	MACD	MOM	
	p-value	**0.00**	0.64	0.76	0.78	
	Variable	RSI	SMA*	WillR	V*	R*
	p-value	0.94	0.71	0.30	**0.01**	0.22

It can be seen that, for CSI 100 stock markets, the structure disagreement is mainly affected by Japanese markets and internal variables of stock markets, which include the Hang Seng index, net capital inflows, and trading volumes. Hong

Kong and Japan, critical economic entities in the Asia-Pacific region, have a significant influence on Chinese stock markets. Moreover, net capital inflows and trading volumes reflect the investors' attention to the stock market, and they also reflect the disagreement of stock markets.

Further, we apply the multiple regression with intercepting terms, where the dependent variable is structure disagreement with stock clusters 5. Explanatory variables only include significant variables, Hang Seng index (HSI*), Nikkei 225 index (N225*), net capital inflows (NCI*), and volumes (V*), in the above single regression model. In order to scientifically explain the market structure disagreement, we use two structure disagreement variables with the clustering characteristic of the log difference of stock daily trading volumes and turnover rate. Considering the colinearity between explanatory variables, we use a stepwise regression method for analysis. The regression results of forwarding and backward models are consistent and shown in Table XVI.

TABLE XVI
THE RESULT OF REGRESSION MODELS WITH MULTIPLE EXPLAIN VARIABLES. ***, **, AND * DENOTE THE SIGNIFICANCE OF t-TEST AT 1%, 5%, AND 10%, RESPECTIVELY. VALUES IN PARENTHESIS ARE STANDARD ERRORS.

Variable	SD(k=5)	
	Daily trading volumes	Daily turnover rate
HSI*	-	0.0619
	-	(0.0730)
N225*	0.1023*	0.0765
	(0.0545)	(0.0723)
V*	0.0113***	0.0093**
	(0.0044)	(0.0043)
NCI*	-0.0001***	-0.0001***
	(0.0000)	(0.0000)
c	0.7412***	0.7412***
	(0.0017)	(0.0017)
Adjust R^2	0.0359	0.0343

From the multiple regression results, we can find that: (1) The structure disagreement of CSI 100 stock markets is significantly affected by the Hang Seng index, Nikkei 225 index and the internal variables of the market, especially trading volumes and net capital inflows. (2) Especially when the log difference of trading volumes increases, the structure disagreement becomes significantly large. (3) However, the influence of the growth rate of net capital inflows is negative on the structure disagreement. Besides, the adjusted R^2 of the regression model is only 0.036, which indicates that the structure disagreement mainly includes the structure information of the stock market and can not be replaced by trading volumes or other variables. In practice, investors should combine trading volumes, net capital inflows and other variables when using the market structure disagreement to predict index returns. In addition, the Chinese stock market is still weak and ineffective because the structure disagreement improves the prediction of index returns.

VI. CONCLUSIONS

The prediction of the stock market is difficult. Some studies showed that disagreement could improve the prediction of stock markets. However, previous researches only used the divergence of investors' opinion or belief, volumes, or turnover rate to measure or replace the disagreement, not considering the market structure.

In this paper, at first, we introduce a new concept of structure disagreement and measure it based on the Kmeans clustering algorithm and the Gini impurity. Then, we analyze the prediction of index returns from the market structure disagreement by machine learning methods, including Linear Regression, AdaBoost, GradientBoosting, XGBoost, and RandomForest. In addition, we analyze the robustness of the predictability of the market structure disagreement. Finally, the influence factors of the market structure disagreement are also studied.

The results of experiments for CSI 100 show that: (1)The market structure disagreement can further improve the predicted direction accuracy of machine learning methods, compared with the trading volume. The average accuracy of prediction models with structure disagreement is about 1.5% higher than that without structure disagreement. In financial practice, this improvement has an important and significant impact on the return of timing investment. (2)The market structure disagreement is mainly affected by the trading volume and net capital inflow. Notably, it increases with the log difference of the trading volume and the decrease with the growth rate of net capital inflows.

In addition, we also find some meaningful results. (1) There are not strong relations between predicted fitness error and direction accuracy of index returns by machine learning methods. (2)Each factor may have its own property, so we can not merely use a basket of factors to predict the movement of indices. (3)The prediction of the CSI 300 market indices is more complicated than that of the CSI100 market indices. (4) It is more suitable for the medium-term forecasting in Chinese stock market, which means investors can not frequently adjust the position. These findings can help investors to further systematically understand the prediction of index returns, and can also help improving the effectiveness of Chinese stock markets.

Our research has the following advantages. (1) It's the first time to research the factor of disagreement from market structure for aggregate stock markets. (2) We propose a method to measure this structure disagreement. (3) We investigate the index prediction from the structure disagreement by machine learning models. Although the improvement is only around 1.5%, the financial market is very difficult to predict, so it can not be ignored. This research further perfects the theory of the disagreement in finance. However, there are also some improvements in the future. We only analyze Chinese stock markets and don't analyze other aggregate stock markets, like S&P 500 and FEST 100 markets. Also, other measurement methods of the structure disagreement are worth exploring. Last but not least, compared to the econometric model, do these relations change? All these issues will be further studied.

REFERENCES

[1] B. G. Malkiel and E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[2] B. G. Malkiel, "Passive investment strategies and efficient markets," *European Financial Management*, vol. 9, no. 1, pp. 1–10, 2003.

[3] S. Richardson, rnIrem Tuna, and rnPeter Wysocki, "Accounting anomalies and fundamental analysis: A review of recent research advances," *Journal of Accounting & Economics*, vol. 50, no. 2–3, pp. 410–454, 2010.

[4] A. W. Lo, H. Mamaysky, and J. Wang, "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation," *Journal of Finance*, vol. 55, no. 4, pp. 1705–1770, 2000.

[5] A. Ang and G. Bekaert, "Stock return predictability: Is it there?" *Review of Financial Studies*, vol. 20, no. 3, pp. 651–707, 2007.

[6] E. Chong, C. Han, and F. C. Park, "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies," *Expert Systems with Applications*, vol. 83, pp. 187–205, 2017.

[7] X. Zhong and D. Enke, "Forecasting daily stock market return using dimensionality reduction," *Expert Systems With Applications*, vol. 67, pp. 126–139, 2017.

[8] E. F. Fama and K. R. French, "The cross section of expected stock returns," *Journal of Finance*, vol. 47, no. 2, pp. 427–465, 1992.

[9] ——, "International tests of a five factor asset pricing model," *Journal of Financial Economics*, vol. 123, no. 3, pp. 441–463, 2017.

[10] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of arima and artificial neural networks models for stock price prediction," *Journal of Applied Mathematics*, vol. 2014, no. 1, pp. 1–7, 2014.

[11] P. R. A. Firmino, P. S. G. D. M. Neto, and T. A. E. Ferreira, "Error modeling approach to improve time series forecasters," *Neurocomputing*, vol. 153, pp. 242–254, 2015.

[12] Y. Kara, M. A. Boyacioglu, and O. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines," *Expert Systems With Applications*, vol. 38, no. 5, pp. 5311–5319, 2011.

[13] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems With Applications*, vol. 42, no. 1, pp. 259–268, 2015.

[14] J. Wang, R. Hou, C. Wang, and L. Shen, "Improved v -support vector regression model based on variable selection and brain storm optimization for stock price forecasting," *Applied Soft Computing*, vol. 49, pp. 164–178, 2016.

[15] L. S. Malagrino, N. T. Roman, and A. M. Monteiro, "Forecasting stock market index daily direction: A bayesian network approach," *Expert Systems With Applications*, vol. 105, pp. 11–22, 2018.

[16] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch type models," *Expert Systems With Applications*, vol. 103, pp. 25–37, 2018.

[17] C. Bilson, T. J. Brailsford, and V. J. Hooper, "Selecting macroeconomic variables as explanatory factors of emerging stock market returns," *Pacific basin Finance Journal*, vol. 9, no. 4, pp. 401–426, 2001.

[18] M. J. Flannery and A. Protopapadakis, "Macroeconomic factors do influence aggregate stock returns," *Review of Financial Studies*, vol. 15, no. 3, pp. 751–782, 2002.

[19] P. M. Dechow, A. P. Hutton, L. K. Meulbroek, and R. G. Sloan, "Short sellers, fundamental analysis, and stock returns," *Journal of Financial Economics*, vol. 61, no. 1, pp. 77–106, 2001.

[20] X. Yan and L. Zheng, "Fundamental analysis and the cross-section of stock returns: A data-mining approach," *Review of Financial Studies*, vol. 30, no. 4, pp. 1382–1423, 2017.

[21] W. Wong, M. Manzur, and B. Chew, "How rewarding is technical analysis? evidence from singapore stock market," *Applied Financial Economics*, vol. 13, no. 7, pp. 543–551, 2003.

[22] C. J. Neely, D. E. Rapach, J. Tu, and G. Zhou, "Forecasting the equity risk premium: The role of technical indicators," *Management Science*, vol. 60, no. 7, pp. 1772–1791, 2014.

[23] H. Hong and J. C. Stein, "Disagreement and the stock market," *Journal of Economic Perspectives*, vol. 21, no. 2, pp. 109–128, 2007.

[24] S. D. Baker, B. Hollifield, and E. Osambela, "Disagreement, speculation, and aggregate investment," *Journal of Financial Economics*, vol. 119, no. 1, pp. 210–225, 2016.

[25] B. I. Carlin, F. A. Longstaff, and K. Matoba, "Disagreement and asset prices," *Journal of Financial Economics*, vol. 114, no. 2, pp. 226–238, 2014.

[26] H. Hong and J. C. Stein, "Differences of opinion, short-sales constraints, and market crashes," *Review of Financial Studies*, vol. 16, no. 2, pp. 487–525, 2003.

[27] E. F. Fama and K. R. French, "Disagreement, tastes, and asset prices," *Journal of Financial Economics*, vol. 83, no. 3, pp. 667–689, 2007.

[28] R. Sadka and A. Scherbina, "Analyst disagreement, mispricing, and liquidity," *Journal of Finance*, vol. 62, no. 5, pp. 2367–2403, 2007.

[29] K. B. Diether, C. J. Malloy, and A. Scherbina, "Differences of opinion and the cross section of stock returns," *Journal of Finance*, vol. 57, no. 5, pp. 2113–2141, 2002.

[30] L. Cen, K. C. J. Wei, and L. Yang, "Disagreement, underreaction, and stock returns," *Management Science*, vol. 63, no. 4, pp. 1214–1231, 2017.

[31] C. Park, "Stock return predictability and the dispersion in earnings forecasts," *Journal of Business*, vol. 78, no. 6, pp. 2351–2376, 2005.

[32] J. Yu, "Disagreement and return predictability of stock portfolios," *Journal of Financial Economics*, vol. 99, no. 1, pp. 162–183, 2011.

[33] M. Baker and J. Wurgler, "Investor sentiment in the stock market," *Journal of Economic Perspectives*, vol. 21, no. 2, pp. 129–151, 2007.

[34] J. S. Kim, D. Ryu, and S. W. Seo, "Investor sentiment and return predictability of disagreement," *Journal of Banking & Finance*, vol. 42, no. 1, pp. 166–178, 2014.

[35] A. Siganos, E. Vagenas Nanos, and P. Verwijmeren, "Divergence of sentiment and stock market trading," *Journal of Banking & Finance*, vol. 78, pp. 130–141, 2017.

[36] S. Basak, "Asset pricing with heterogeneous beliefs," *Journal of Banking & Finance*, vol. 29, no. 11, pp. 2849–2881, 2005.

[37] A. Beber, F. Breedon, and A. Buraschi, "Differences in beliefs and currency risk premiums," *Journal of Financial Economics*, vol. 98, no. 3, pp. 415–438, 2010.

[38] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna, "Cluster analysis for portfolio optimization," *Journal of Economic Dynamics & Control*, vol. 32, no. 1, pp. 235–258, 2008.

[39] B. B. Nair, P. K. S. Kumar, N. R. Sakthivel, and U. Vipin, "Clustering stock price time series data to generate stock trading recommendations: An empirical study," *Expert Systems with Applications*, vol. 70, pp. 20–36, 2017.

[40] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[41] T. W. Liao, "Clustering of time series data-a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[42] S. Nanda, B. Mahanty, and M. Tiwari, "Clustering indian stock market data for portfolio management," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8793–8798, 2010.

[43] A. W. Rosyadi and N. Suciati, "Image segmentation using transition region and k-means clustering," *IAENG International Journal of Computer Science*, vol. 47, no. 1, pp. 47–55, 2020.

[44] D. Li, D. Yang, J. Zhang, and X. Zhang, "Ar-ann: Incorporating association rule mining in artificial neural network for thyroid disease knowledge discovery and diagnosis," *IAENG International Journal of Computer Science*, vol. 47, no. 1, pp. 25–36, 2020.

[45] Y. Sun, Y. Chen, Y. Pan, and L. Wu, "Android malware family classification based on deep learning of code images," *IAENG International Journal of Computer Science*, vol. 46, no. 4, pp. 524–533, 2019.

[46] N. Shigei, K. Mandai, S. Sugimoto, R. Takaesu, and Y. Ishizuka, "Land-use classification using convolutional neural network with bagging and reduced categories," in *In Proceedings of The International MultiConference of Engineers and Computer Scientists*, Hong Kong, 2019, pp. 7–11.

[47] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. WILEY, 2013.

[48] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the Second European Conference on Computational Learning Theory*, 1995.

[49] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[50] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[51] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[52] J. Han and Z. Ge, "Effect of dimensionality reduction on stock selection with cluster analysis in different market situations," *Expert Systems With Applications*, vol. 147, p. 113226, 2020.

[53] Y. Zhang and J. Yao, "Gini objective functions for three-way classifications," *International Journal of Approximate Reasoning*, vol. 81, pp. 103 – 114, 2017.

[54] E. Muchai and L. Odongo, "Comparison of crisp and fuzzy classification trees using gini index impurity measure on simulated data," *European Scientific Journal, ESJ*, vol. 14, no. 18, p. 351, 2018.