6. A MULTI-INDICATOR MULTI-OUTPUT MIXED FREQUENCY SAMPLING APPROACH FOR STOCK INDEX FORECASTING

Yuchen PAN[1]
Zhi XIAO[2]
Xianning WANG[3]
Daoli YANG[4]

Abstract

Compared to stock price index sampled at higher frequency, its indicators are usually sampled at lower frequencies. In practice, with higher frequency variables response to lower frequency variables, we can get multi-output for each period. This paper explores how to construct a multi-indicator multi-output (MIMO) mixed sampling frequency approach for stock price index forecasting. We also consider nonlinear relationship between dependent variables and independent variables in stock market. We establish a new model by applying multiple output support vector machine (MSVM) to modify mixed data sampling (MIDAS). We compare results with other models and make DM tests. The experiment shows that the proposed model performances better.

Keyword: Stock price index forecasting, MIMO, MIDAS, MSVM, nonlinearity

JEL Classification: C53, E47

[1] *Corresponding Author. Business School at Southwest University of Political Science&Law. No. 399 Baosheng Road, Yubei, Chongqing, 401120, China. Email:pan0211yuchen@sina.com*

[2] *Corresponding Author, School of Economics and Business Administration at Chongqing University. No. 174 Shazhengjie, Shapingba, Chongqing, 400044, China. Email: xiaozhicqu@163.com*

[3] *School of Economics and Management at Chongqing Normal University. No.37, Middle Road of University Town, Shaping District, Chongqing, 401331, China; Big Data Marketing Research and Applications Center of Chongqing Normal University, No. 37, Middle Road of University Town, Shapingba District, Chongqing, 401331, China.Email: xianningwang@sina.com*

[4] *E-commerce Department of School of Business Planning at Chongqing Technology and Business University. No. 19, Xuefu Avenue, Nan'an, Chongqing, 400067, China*

1. Introduction

Stock prices forecasting are always interesting, and researchers continue studying it from different perspectives and approaches (Lupu, 2015; Wang, 2014). In stock price index forecasting, we often face data sampled at different frequencies. For instance, stock prices index are usually sampled at daily or weekly, but indicators like financial index or macroeconomic factors are sampled at monthly or quarterly. Thus, researchers now focus on mixed frequency sampling objectively existed in stock market. Researchers in tradition addressed this issue by averaging lower frequency data or aggregate higher frequency data. Ghysels (2004) argued that the motivation to propose a mixed frequency approach is adopting original data rather than preprocessing them usually improves forecasting performances. Researchers later prove that mixed sampling frequency approaches have strong ability in explaining real cases in stock market in different countries.

Besides, indicators like financial index or macroeconomic factors are sampled at lower frequency than stock price index. When lower frequency indicators forecast higher frequency dependent variables, there exist multi-outputs. For example, when a monthly CPI forecasts weekly stock index, the monthly CPI has impact on each of the four week of stock index with different degrees. Each week of stock index in that month has relations to each other. This is a multi-output issue. Researches have proved multi-dimensional outputs (Platt, 1999; Weston, 1999; Crammer, 2001; Nedaie, 2016) and time-series outputs (Suykens, 1999) exist in practical occasions. Our case belongs to the later one. The biggest advantage of multi-output approach is to focus on overall impact on sequential outputs, extend prediction range with limited known data information, and inherently preserve dependencies among prediction results (Aho, 2012; Blockeel, 1999; Breiman, 1997; Kocev,2009). This in turn helps to explain real case with stronger ability (Aho, 2012; Blockeel, 1999; Suzuki, 2001; Ženko, 2008). These potential dependencies are objectively existed in stock index forecasting and many other economic cases. Single output approaches ignore the dependencies among outputs. Another advantage of multi-output approach, especially for case of time-series outputs, is to be able to recognize both short-term pattern and long-term pattern, and ultimately demonstrate the trend of time series occasion (Caruana, 1997). Time-series forecasting is very popular and important in stock index forecasting. Single output approach, however, is difficult to recognize long-term patterns. Based on the above two advantages, adopting multi-output approach obtains better results than single-output approach does (Crammer, 2002; Platt, 2000; Suykens, 1999; Weston, 1999). Next, multi-output approach is more flexible in forecasting (Sánchez-Fernández, 2004; Tuia, 2011).

For statistical and announcement systems, many kinds of data are offered recently that leads to limited sample size. At the same time, data is updated quickly with strong timeliness nowadays. Recent data which has only small sample size holds stronger explain ability in practice. When confronted with this situation, researchers explore corresponding approaches to make full use data with small sample size.

Our interests include the above three aspects. We propose a nonlinear MIMO-MIDAS (MIxed Data Sampling) model to explore multi-indicator multi-output (MIMO) relationship in stock market using small sample size. Experts and economists concluded that stock prices are fluctuating in a highly nonlinear and dynamic way (Hiemstra and Jones, 1994). MIDAS approach cannot directly address nonlinearity and small sample size. We modify U-MIDAS approach by multiple support vector machine (MSVM), which owns the ability to tackle multiple outputs (Mao et al., 2014; Han, 2012; Tuia, 2011; Sánchez-Fernández, 2004; Xiong *et al*., 2014), nonlinearity (Chang, 2015; Cheng, 2013) and small sample size manually according to its inherent merit. This is an implicit advantage of our model. In the experimental part, we apply stock prices index in China from 2008 to 2017. We compare our model with other models and show better performances over the benchmarks.

The rest of the paper is structured as follows. Section 2 reviews previous literatures. Section 3 introduces MIDAS, U-MIDAS and MSVM approach respectively. Section 4 builds our novel model. Section 5 describes data and measurements. Section 6 shows results. Section 7 concludes.

2. Literature Review

2.1 MIDAS

Mixed data sampling (MIDAS) is now the most popular approach for mixed frequency data. Researchers investigate MIDAS model from both modification of the model and application in different areas.

Bessec et. al. (2019) introduces a Markov-switching model with mixed data frequency (MSV-MIDAS) to forecast business cycle turning points in the United States. Andreou et. al. (2019) introduces mixed-frequency group factor model to address the issue that whether Industrial Production is the dominant factor in driving the US economy. Asgharian et al.'s research (2013) applied GARCH-MIDAS to explain stock return variance (daily) by macroeconomic variables (monthly). Galvao (2013) explored nonlinearity of MIDAS approach. The latest researches proposed unrestricted MIDAS (U-MIDAS). Xu applies U-MIDAS model into artificial neural network and develops an ANN-U-MIDAS model. Hepenstrick applies mixed data frequency into the three-pass regression fliter to forecast GDP in the US. Foroni et al., (2015) examined performances of both models, and concluded that the U-MIDAS performs even better than traditional MIDAS when the differences between frequencies are small. Barsoun and Stankiewicz (2015) modified U-MIDAS with switching regimes and forecasted GDP growth in different business cycle patterns. MIDAS and U-MIDAS are the most popular methods for mixed frequency sampling and integrate with other methods to address issues in different areas (Wang, 2017). All studies about mixed sampling frequency are for single output.

Many researchers apply mixed frequency data into different areas. Some of them adopted MIDAS to solve mixed frequency sampling issues in stock markets. Ghysels (2005) tested relationship between conditional variance and conditional mean of stock return in trade-off. Ghysels et al. (2006) evaluated various dimensions and horizons impacted on volatility. Forsberg et al. (2007) evaluated measurements of absolute returns volatility. Leon (2007) adopted monthly excess return and the square of the daily excess return to study on relationship between expected return and risk on stock markets of several European countries. Adjaoute (2010) adopted weekly stock index volatility and daily stock return to study the relationship on emerging market. Researchers also applied MIDAS model into other areas. Tsui et.al. (2018) forecasts the Singapore GDP using MIDAS model. Duarte et. al. (2017) used MIDAS model to forecast quarterly private consumption by high frequency data from automated teller machines (ATM) and points-of-sale(POS). Ghysels et.al. (2015) constructed a combination of MIDAS-type regression models to forecast the annual US federal government current expenditures and receipts. Christiane (2015) adopted daily financial data to forecast monthly oil price. Chen and Ghysels (2011) studied how good news and bad news have effect on volatility.

2.2 MSVM

MSVM has its roots in single support vector machine (SVM), which is widely used in stock market (Thenmozhi (2016), Żbikowski (2015)). SVM was proposed by Vapnik(1995). It requires minimal structural risk and globally optimized solutions (Vapnik (1998)). This characteristic helps SVM have special advantage on small sample size because structural risk minimization minimizes the requirement for sample size. The basic principle is to find a separator hyperplane through nonlinear mapping and to construct an ε-insensitive tube in that feature space with a maximum margin to classify data samples into two classes. This principal realizes nonlinearity which makes up the drawback of MIDAS.

SVM cannot address multiple output issues. Perez-Cruz (2002) proposed the efficient multiple SVM extended from SVM. Multiple SVM is also called multidimensional regression problem. This problem can be regarded as multiple one-dimensional problems, which can be solved by dimension reduction method. It transforms that multi-dimensional output problem into man one-dimensional output problems. However, when original SVM for regression estimation is applied over each direction of a multi-dimensional problem, samples will not be equally penalized (Perez-Cruz, 2002). It ignores the correlation between the dimensions of the output data and finally reduces the accuracy of the model.

Many researchers explore MSVM from both theory and methods. Weston (1999) and Vapnik(1998) have done similar research to extend the two classification problems to the multi classification problems, aiming to improve the classification ability, but they both limit the classifier. Platt (1999) proposed the Decision Directed Acyclic Graph, to combine many two-class classifiers into a multiclass classifier. Perez-Cruz (2001) proposed iterative reweighted least square method to instead the previous optimization method of MSVM, which has become a common method of MSVM. Later he proposed how to speed up its algorithm (Perez-Cruz 2004)。Crammer (2002) discusses how to transform the two classification problem into the multi classification problem without restriction, so as to avoid the quadratic homogeneous problem. Mao *et al*.(2014a) applied LOO（virtual leave-one-out）error estimation method to MSVM to instead the traditional LOO error estimation, and reduced operation time and cost through cross validation procedure. After that, they proposed a new model selection algorithm to improve the operation time and stability (Mao et al., 2014b). Bao (2014) applies MSVM with MIMO strategy to realize multi-step-ahead forecasting. Previous researches indicate that MSVM obtains better predictions than single output SVM independently for each dimension (Perez-Cruz, 2001, 2002; Sanchez-Fernandez, 2004; Tuia, 2011; Han, 2012), and generally improves performance especially for small sample size. MSVM then becomes one of the most widely used methods to solve the multiple output problems. It is widely used in different fields (Perez-Cruz, 2001, 2002; Sanchez-Fernandez, 2004; Tuia, 2011; Mao, 2014b). In economics, Xiong et al. (2014) proposed MSVM for interval time series forecasting to value the stock price index. Cherchye et. al. (2016, 2014) researched on the impact of cost efficiency on multi-output profit setting.

3. Related Methods

To realize MIMO-MIDAS, we adopt U-MIDAS with MSVM methods in our paper. We briefly review the two methods in the following contents.

3.1 The MIDAS approach

MIDAS approach was proposed by Ghysels et al. (2004, 2006) that allows indicators and outputs sampled at different frequencies. It is a time-series regression model and origins from distributed lag polynomials. MIDAS approach is parameterized in a flexible function to run parsimoniously, and can be straightly extended to multiple indicators. The basic MIDAS approach for a single independent variable with h-step ahead is given by:

$$y_t = \beta_0 + \beta_1 B(L^{1/m};\theta)x_{t-h}^{(m)} + \varepsilon_t \tag{1}$$

Where $B(L^{1/m};\theta) = \sum_{k=0}^{K} b(k;\theta)L^{k/m}$, and sums the weights. K represents lags of the independent variables (higher frequency variables). $b(k;\theta)$ is the kth weight of K -lag polynomial, calculated by a certain function of θ parameters. L is the lag operator, and $L^{1/m}$ operates at the higher frequency. Here t represents the basic time unit, and m is the higher sampling frequency. $L^{k/m}x_t^{(m)} = x_{t-k/m}^{(m)}$. All parameters in the Midas model are obtained according to the h period. h emphasizes which period to start data acquisition and forecast. For example, when the weekly independent variable if used to forecast monthly dependent variable, $m = 4$. When $h = 3/4$, The first quarter of the data information of the current month was obtained and used. While when$h = 1/4$, the first three weeks' data information of the current month was obtained forforecasting.

Exponential Almon function is the most popular weight functions used in MIDAS regressions. It usually uses two parameters in MIDAS approach for simplicity, but keeps flexibility in the specification. The $b(k;\theta)$ is parameterized as

$$b(k;\theta) = \frac{exp(\theta_1 k + \theta_1 k^2)}{\sum_{k=1}^{K} exp(\theta_1 k + \theta_1 k^2)} \tag{2}$$

This specification ensures that weights are positive and add up to one. It also produces a wide variety of shapes for different values of the two parameters. The parameter estimation is done by non-linear least squares method.

3.1.1 Multiple Leading Indicator Models

The basic MIDAS approach can be directly extended to multi-indicator specification without increasing a large number of parameters needed to be estimated. The multi-indicator model would be:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_{1i} B_i(L^{1/m};\theta_i)x_{i,t-h}^{(m)} + \varepsilon_t \tag{3}$$

where the multiple indicators are numbered by $i = 1, \ldots n$. β_{1i} parameters define the weights attached to indicators, and are specific to the forecast horizon. Each indicator needs the estimation of only two parameters to confirm the lag structure (θ_i), and one to weight its impact on $y_t(\beta_{1i})$.

The multiple indicators autoregressive MIDAS model with h-steps-ahead would be written as:

$$y_t = \beta_0 + \gamma y_{t-h} \sum_{i=1}^{n} \beta_{1i} B_i(L^{1/m};\theta_i)x_{i,t-h}^{(m)} + \varepsilon_t \tag{4}$$

3.2 The Unrestricted MIDAS

Foroni et al. (2015) proposed a novel MIDAS model without restrictions on the weights of lag polynomial, and named it unrestricted MIDAS (U-MIDAS):

$$y_t = \beta_0 + \sum_{j=0}^{J} \beta_{j+1} x_{t-h-j/m}^{(m)} + \varepsilon_t$$

The multi-indicator U-MIDAS is defined by:

$$y_t = \beta_0 + \sum_{i=1}^{n} \sum_{j=0}^{J} \beta_{i,j+1} x_{i,t-h-j/m}^{(m)} + \varepsilon_t \quad (5)$$

This notation is consistent with Eq. (1). However, no structure is imposed on the shape of weights of lag polynomial. All parameters in Eq. (5) are needed to be estimated. Foroni et al. (2015) and Barsoun and Stankiewicz (2015)) indicated that if the difference in frequencies between indicators and outputs is small, the parameter proliferation in U-MIDAS approach is not problematic for multiple indicators. U-MIDAS approach can be estimated by ordinary least squares, which can simplify the calculation.

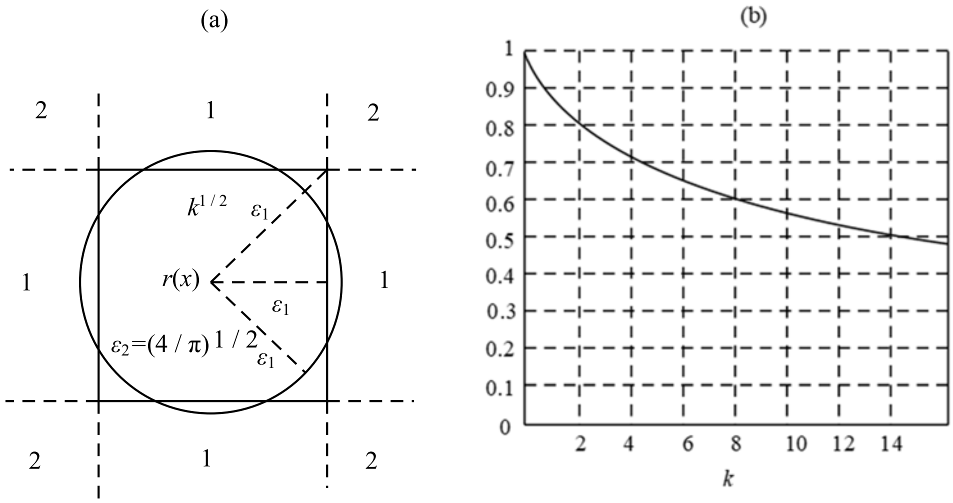
3.3 The MSVM model

The MSVM mainly improves the algorithm of loss function in the SVM. The loss function defined on the hypersphere replaces that defined on the hypercube. When the hyper-spherical (instead of hypercube in SVM) insensitive zone is defined in MSVM, it is able to treat every sample equally (Perez-Cruz, 2002). In this way, it enhances the correlation between the output components. Samples are penalized by the same factor in this insensitive zone. In Figure 1, graph (a) represents a two-dimensional regression problem. The square and the circle represent the insensitive region of the hypercube insensitive region and the hyper-sphere, respectively. They have the same area. To show that the hypercube and hyperplane have the same super capacity, we show the scale of ε_1 and ε_2 in graph (b). SVM defines an insensitive region for regression estimation. In the process of parameter estimation, if the distance between a sample and its estimated value exceeds ε (the insensitive area in the definition), the penalty factor C will punish it. If the existing one-dimensional SVM is applied to multi-dimensional problems directly, it may lead to two problems. First, some samples located in the insensitive area will be wrongly judged as located out of the insensitive area, which will lead to punishment. Second, it may lead to the repeated addition of penalty factors due to a certain sample. For example, a sample whose difference between each dimension and the estimated value exceeds ε will cause the penalty factor C to be added k times (k is the dimension).

The redefined insensitive loss function evaluates the risk of multiple output components at the same time. The new loss function can consider the fitting error of each component simultaneously, so that the objective function is related to the error of each component, which achieves an overall optimized goal. It also weakens the influence of noise on the result and improve the anti-noise performance of the algorithm. Finally, we adopt the iterative reweighted least square method to solve the corresponding dual problem, and establish a multi output model. It enables to establish a high-precision multi output regression model by selecting the appropriate kernel function and related parameters.

Figure 1

Hyper-cubic and hyper-spherical insensitive zone



The detailed MSVM can be found in Perez-Cruz et al. (2002). A brief introduction is demonstrated as follows.

Suppose the observable output is a vector with Q variables, i.e.,y ∈ R^Q. Given a labeled training data set $((x_i, y_i)\forall i = 1, \ldots n$, where $x_i \in R^d$ and $y_i \in R^Q$) and a mapping function which is a nonlinear transformation to a higher dimensional space (φ(·),$R^d \xrightarrow{\varphi(\cdot)} R^Q$ and $d \leq Q$). The formulation is as follows:

$$\min Lp(\mathbf{w}, \mathbf{b}) = \frac{1}{2}\sum_{j=1}^{Q}\|w^j\|^2 + C\sum_{i=1}^{n} L(u_i) \quad (6)$$

where $\mathbf{W}$=[$\mathbf{w}^1, \ldots, \mathbf{w}^Q$] and $\mathbf{b}$=[$\boldsymbol{b}^1, \ldots, \boldsymbol{b}^Q$]^T define k-dimensional linear regressor in the Q-dimensional feature space. C is a hyper parameter, and determines the trade-off between the regularization and the error reduction term.

$$u_i = \|\mathbf{e}_i\| = \sqrt{\mathbf{e}_i{}^T\mathbf{e}_i} \quad (7)$$

$$\mathbf{e}_i{}^T = \mathbf{y}_i{}^T - \boldsymbol{\varphi}^T(\mathrm{x}_i)\mathbf{W} - \mathbf{b}^T \quad (8)$$

$\boldsymbol{\varphi}(\cdot)$represents mapping function from primal space to a higher dimensional feature space. $L(u_i)$ denotes a quadratic ε-insensitive cost function.

$$L(\mathrm{u}) = \begin{cases} 0, & u < \varepsilon \\ u^2 - 2u\varepsilon + \varepsilon^2 & u \gg \varepsilon \end{cases} \quad (9)$$

This is where the difference is compared with that of Vapnik ε-insensitive loss function. The cost function $L(\mathrm{u})$ enables MSVM capable of finding dependencies between outputs, and taking advantage of information of all outputs to get a robust solution. In detail, when ε is nonzero, it considers all outputs to construct each individual regressor and obtains more stable predications. Then it yields a single support vector set for all dimensions. Since Eq. (6) cannot be solved straightforwardly, Eq. (9) adopted iterative re-weighted least squares (IRWLS) to obtain a desired solution. By introducing a first-order Talor expansion of the cost function $L(\mathrm{u})$, the objective of Eq. (6) will be approximated by the following equation (Tuia et al., 2011):

$$Lp'(\mathbf{W},\mathbf{b}) = \frac{1}{2}\sum_{j=1}^{Q}\|w^j\|^2 + \frac{1}{2}\sum_{i=1}^{n}\alpha_i u_i^{\,2} + CT) \tag{10}$$

where

$$\alpha_i = \begin{cases} 0, & u_i^k < \varepsilon \\ \frac{2C(u_i^k-\varepsilon)}{u_i^k} & u_i^k \gg \varepsilon \end{cases} \tag{11}$$

$CT$is constant term which does not depend on $\mathbf{W}$ and $\mathbf{b}$, the superscript k represents kth iteration.

To optimize Eq. (10), the IRWLS procedure is constructed, and linearly searched the next step solution along descending direction based on previous solutions. According to Representer Theorem (Schölkopf and Smola, 2002), the best solution of minimization of Eq. (10) in feature space can be expressed as $w^j = \sum_i \varphi(\mathrm{x}_i)\beta^j = \boldsymbol{\varphi}^T\beta^j$, so the goal is transformed to finding optimal β and b. The IRWLS of MSVM can be summarized in the following steps (Sánchez-Fernández et al., 2004):

Step 1: Initialization: Set k=0 , $\beta^k = 0$, and $b^k = 0$. Calculate u_i^k and α_i.

Step 2: Compute the solution β^s and b^s according to the nest equation:

$$\begin{bmatrix} \mathbf{K}+\mathbf{D}_\alpha^{-1} & 1 \\ \boldsymbol{\alpha}^T\mathbf{K} & \mathbf{I}^T\boldsymbol{\alpha} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta}^j \\ \mathrm{b}^j \end{bmatrix} = \begin{bmatrix} y^j \\ \boldsymbol{\alpha}^T y^j \end{bmatrix} \tag{12}$$

where $\boldsymbol{\alpha}$=$[\alpha_1, \dots \alpha_l]^T$, for $j = 1, \dots Q$, $(\mathbf{D}_{\boldsymbol{\alpha}})_{ij} = \alpha_i\delta(i-j)$, $\mathbf{I}$ is a column vector of n ones, $\mathbf{K}$ is the kernel matrix, and $\mathbf{K}(x_i \cdot x_j) = (\varphi(x_i)\cdot\varphi(x_j))$. Define the corresponding descending direction $P^k = \begin{bmatrix} w^s - w^k \\ (\mathrm{b}^s-\mathrm{b}^k)^T \end{bmatrix}$.

Step 3: Use a backtracking algorithm to compute $\boldsymbol{\beta}^{k+1}$ and $\boldsymbol{b}^j$.

Step 4: Obtain u_i^{k+1} and α_i. Go back to Step 2 until convergence.

The proof of convergence of the above algorithm is given in previous study (see example in (Sánchez-Fernández et al., 2004). Once convergence is reached, $\boldsymbol{\beta}^{k+1}$ and $\boldsymbol{b}^j$ are parameters of jth output regressor. Since u_i^k and α_i are calculated using every dimension of y, each individual regressor contains all outputs information, which improves the prediction performance of MSVM.

4. The proposed model

4.1 Model Construction

This paper explains higher frequency dependent variables with lower frequency independent variables. The univariate basic model is：

$$y_{tq} = \alpha_q + \sum_{j=0}^{J} \beta_{q,(j+1)} (x_q)_{t-h-j/m}^{m} + \varepsilon_{tq}$$

This is the qth output. α_q is the constant. t represents the basic time unit, which is "one week" in our case. The time unit of qth and $(q+1)$th output is consistent with t. m is times between the dependent variable and independent variable. h is the beginning of lags recursed by week. j represents the lag, and is also recursed by week. ε_{tq} is the error.

We generate low-frequency data into high-frequency data. q and t represent the same time unit, so $(x_q)_{t-h-j/m}^{m}$ represents the lower frequency data value(monthly data in our case) correspondent to time point $(q+t-h-j/m)$. That is, the lower frequency data value within m. Values of certain weeks are affected by those of same months, but their values are varying by time. Thus, $\beta_{q,(j+1)}$ is the coefficient of $(x_q)_{t-h-j/m}^{m}$, and varies from week to week. Simultaneously, we recurse with existing monthly data value by weighted combination and coefficient correction to reflect dynamic changes.

The multi-indicator multi-output AR model with h-step-ahead is as follows:

$$\begin{cases} y_{t1} = \alpha_1 + \sum_{i=0}^{I} \gamma_{1,i+1} y_{1,t-h'-i/m} + \sum_{p=1}^{P} \sum_{j=0}^{J} \beta_{1,p(j+1)} (x_1)_{p,t-h-j/m}^{m} + \varepsilon_{t1} \\ y_{t2} = \alpha_2 + \sum_{i=0}^{I} \gamma_{2,i+1} y_{2,t-h'-i/m} + \sum_{p=1}^{P} \sum_{j=0}^{J} \beta_{2,p(j+1)} (x_2)_{p,t-h-j/m}^{m} + \varepsilon_{t2} \\ \qquad \cdots \\ y_{tq} = \alpha_q + \sum_{i=0}^{I} \gamma_{q,i+1} y_{q,t-h'-i/m} + \sum_{p=1}^{P} \sum_{j=0}^{J} \beta_{q,p(j+1)} (x_q)_{p,t-h-j/m}^{m} + \varepsilon_{tq} \\ \qquad \cdots \\ y_{tQ} = \alpha_Q + \sum_{i=0}^{I} \gamma_{Q,i+1} y_{Q,t-h'-i/m} + \sum_{p=1}^{P} \sum_{j=0}^{J} \beta_{Q,p(j+1)} (x_Q)_{p,t-h-j/m}^{m} + \varepsilon_{tQ} \end{cases} \qquad (13)$$

$q = 1,2,\ldots,Q$, representing qth output and there are Q outputs in total. i represents the lag for y, which is recursed by week. h' is the beginning of lags for y_q. $\gamma_{q,i+1}$ is the coefficient of $y_{q,t-h'-i/m}$, and varies from week to week. We apply the sum $\sum_{i=0}^{I} \gamma_{q,i+1}\, y_{q,t-h'-i/m}$ because we recurse with existing weekly data value to reflect dynamic changes. p represents number of indicators, and $p = 1, 2, \ldots, P$.

We extend and convert the above formula to an easier and simpler one:

$$\begin{cases} y_{t1} = \alpha_1 + (\gamma_{1,1}, \dots, \gamma_{1,I+1})(y_{1,t-h'}, \dots, y_{1,t-h'-i/m})^T \\ +(\beta_{1,11}, \dots, \beta_{1,1(J+1),\dots,}\beta_{1,P1}, \dots, \beta_{1,P(J+1)}) \cdot \\ \left((x_1)^m_{1,t-h}, \dots, (x_1)^m_{1,t-h-J/m}, \dots, (x_1)^m_{P,t-h}, \dots, (x_1)^m_{P,t-h-J/m}\right)^T + \varepsilon_{t1} \\ \dots \\ y_{tq} = \alpha_q + (\gamma_{q,1}, \dots, \gamma_{q,I+1})(y_{q,t-h'}, \dots, y_{q,t-h'-i/m})^T \\ +(\beta_{q,11}, \dots, \beta_{q,1(J+1),\dots,}\beta_{q,P1}, \dots, \beta_{q,P(J+1)}) \cdot \\ \left((x_q)^m_{1,t-h}, \dots, (x_q)^m_{1,t-h-J/m}, \dots, (x_q)^m_{P,t-h}, \dots, (x_q)^m_{P,t-h-J/m}\right)^T + \varepsilon_{tq} \\ \dots \\ y_{tQ} = \alpha_Q + (\gamma_{Q,1}, \dots, \gamma_{Q,I+1})(y_{Q,t-h'}, \dots, y_{Q,t-h'-i/m})^T \\ +(\beta_{Q,11}, \dots, \beta_{Q,1(J+1),\dots,}\beta_{Q,P1}, \dots, \beta_{Q,P(J+1)}) \cdot \\ \left((x_Q)^m_{1,t-h}, \dots, (x_Q)^m_{1,t-h-J/m}, \dots, (x_Q)^m_{P,t-h}, \dots, (x_Q)^m_{P,t-h-J/m}\right)^T + \varepsilon_{tQ} \end{cases}$$

$$\begin{cases} y_{t1} = \alpha_1 + \omega_1 \boldsymbol{x}_{1,t-h} + \varepsilon_{t1} \\ y_{t2} = \alpha_2 + \omega_2 \boldsymbol{x}_{2,t-h} + \varepsilon_{t2} \\ \dots \\ y_{tq} = \alpha_q + \omega_q \boldsymbol{x}_{q,t-h} + \varepsilon_{tq} \\ \dots \\ y_{tQ} = \alpha_Q + \omega_Q \boldsymbol{x}_{Q,t-h} + \varepsilon_{tQ} \end{cases}$$

$$\boldsymbol{x}_{q,t-h} = \left(y_{q,t-h'}, \dots, y_{q,t-h'-I/m}, (x_q)^m_{1,t-h}, \dots, (x_q)^m_{1,t-h-J/m}, \dots (x_q)^m_{P,t-h}, \dots, (x_q)^m_{P,t-h-J/m}\right)^T$$

$$\begin{gathered} \omega_1 = (\gamma_{1,1}, \dots, \gamma_{1,I+1}, \beta_{1,11}, \dots, \beta_{1,(J+1)}, \dots, \beta_{1,P1}, \dots, \beta_{1,P(J+1)}) \\ \omega_2 = (\gamma_{2,1}, \dots, \gamma_{2,I+1}, \beta_{2,11}, \dots, \beta_{2,(J+1)}, \dots, \beta_{2,P1}, \dots, \beta_{2,P(J+1)}) \\ \dots \\ \omega_q = (\gamma_{q,1}, \dots, \gamma_{q,I+1}, \beta_{q,11}, \dots, \beta_{q,(J+1)}, \dots, \beta_{q,P1}, \dots, \beta_{q,P(J+1)}) \\ \dots \\ \omega_Q = (\gamma_{Q,1}, \dots, \gamma_{Q,I+1}, \beta_{Q,11}, \dots, \beta_{Q,(J+1)}, \dots, \beta_{Q,P1}, \dots, \beta_{Q,P(J+1)}) \end{gathered}$$

$$\boldsymbol{y}_t = \alpha_{all} + \boldsymbol{\omega} \boldsymbol{X}^m_{MI} + \varepsilon_{all} \tag{14}$$

$\boldsymbol{\alpha}_{all} = [\alpha_1, \dots, \alpha_Q]^T$, $\boldsymbol{\omega} = [\omega_1, \dots, \omega_Q]^T$, $\boldsymbol{X}^m_{MI} = [x_{1,t-h}, \boldsymbol{x}_{2,t-h} \dots, \boldsymbol{x}_{Q,t-h}]^T$, MI represents multi-indicator.

Map $\boldsymbol{X}^m_{MI}$ to high feature space:

$$\boldsymbol{y}_t = \alpha_{all} + \varphi^T(\boldsymbol{X}_{MI}^m)\boldsymbol{\omega} + \varepsilon_{all}$$

According to MSVM model in Section 3.3, the best solution of minimization of Eq. (10) in feature space can be expressed as $\omega = \sum_i \varphi(x_i)\beta^j = \varphi^T\beta^j$, which then is put into Eq.(14):

$$\boldsymbol{Y}_t = \boldsymbol{\alpha}_{all} + \boldsymbol{\varphi}^T(\boldsymbol{X}_{MI}^m)\boldsymbol{\varphi}^T\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{all}$$

$\beta = [\beta_1, \dots, \beta_Q]$, $\boldsymbol{Y_t} = [y_{t1}, \dots, y_{tQ}]^T$, representing Q outputs. $\boldsymbol{\alpha}_{all}$ and $\boldsymbol{\varepsilon}_{all}$ are the whole constant and error including all constants and errors for each equations in (13), respectively.

Our final MIMO-MIDAS model is:

$$\boldsymbol{Y}_t = \boldsymbol{\alpha}_{all} + \mathbf{K}_{\boldsymbol{X}_{MI}^m}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{all} \quad (15)$$

$\mathbf{K}_{\boldsymbol{X}_{MI}^m}$ is the kernel function, and $\boldsymbol{K}\left(\boldsymbol{x_p} \cdot (\boldsymbol{x_q})_{\boldsymbol{p},\boldsymbol{t}-\boldsymbol{h}-\frac{J}{m}}^{\boldsymbol{m}}\right) = \boldsymbol{\varphi^T}(\boldsymbol{x_p}) \cdot \boldsymbol{\varphi}\left((\boldsymbol{x_q})_{\boldsymbol{p},\boldsymbol{t}-\boldsymbol{h}-\frac{J}{m}}^{\boldsymbol{m}}\right)$. Kernel function is a symmetric function, which maps sample points into a higher dimensional feature space, and transforms a nonlinear problem into a linear mode. In detail, it is used to calculate similarity of each pair of objects in input sets. φ denotes the input domain where $\varphi{:}\mathrm{X} \to F$.

Linear, polynomial and radial basis function (RBF) and sigmoid are four main kernel functions. Our study applies RBF because it is the most common one used in previous studies for its well performance and similarity in calculation in most forecasting cases. According to previous literature, RBF is defined as $\emptyset(\tilde{x}) = \exp(\frac{-\|\widetilde{x_i - x_j}\|^2}{2\sigma^2})$. σ is the width of RBF function.

The final performance of MSVM is related to three parameters: regularization constant C, loss function ε, and σ (the width of RBF function). There are no general rules for setting them. We adopt Genetic Algorithm (GA) to search optimal parameters. The basic principle of GA is with simulation of biological evolution phenomenon, the parameter with higher fitness function value is left. A big advantage of GA is to avoid partial minimization and reduce search time. We use GA to select optimal model parameters, locate approximate optimal solution and reduce forecasting errors.

Our proposed model is realized by the following steps.

4.2 Experiment Design

Step 1. Collect and split the data into two sets. The last 10 sample points are for evaluation and the previous of them are for training set. When we get the sampling points, we first standardize our data to ensure that every data point falls into the same parameter range, otherwise some data points may heavily overwhelm others, and errors may be increased.

Step 2. We use GA to search the parameters by fitness function to confirm optimal parameters (C, loss function ε, and σ). The fitness function is:

$$y = \min PER$$

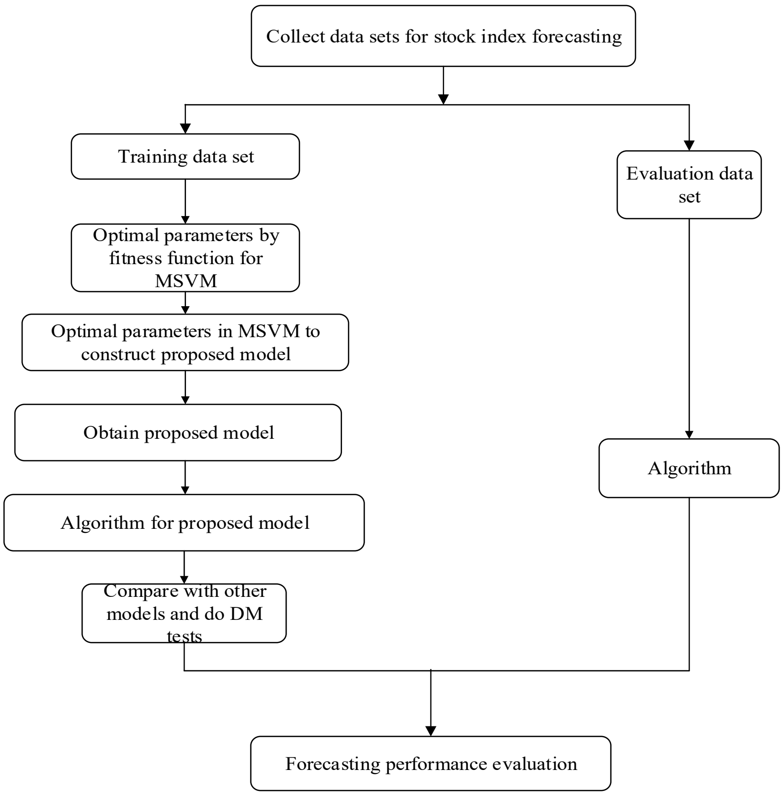
$$PER = \frac{|T_i - P_i|}{T_i}$$

Matlab software package (2014(a)) and LibSVM (version 1.7) are adopted to obtain the final results.

Step 3. According to optimal parameters, we obtain coefficients to construct final MIMO-MIDAS model, compare with other models, do DM tests and get our final results.

Figure 2

The framework of the experiment



5. Results and Performance

5.1. Sample and Data

We choose four weeks' closing index in a month as dependent variables. The data points of stock price index are provided by Shanghai Stock Exchange Composite Index (SSECI). We also choose four macroeconomic indicators. The consumer price index (CPI), the monthly purchasing managers' index (PMI), the new credit and the interest rate. CPI and PMI are obtained from National Bureau of Statistics of the People's Republic of China. The new credit and interest rate are obtained from the People's Bank of China. These indicators are commonly used and proved to be useful in stock price index forecasting. All data are collected from April, 2008 to March, 2017. There are totally 114 sample points in the dataset.

The first 104 sample points are used as the training set, and the remaining data points are used as evaluation points for out of sample forecasting.

Our indicators are available at monthly frequency, while dependent variables are sampled at weekly frequency. Some months contain five weeks but others have only four. In order to have consistent data forms and balanced weekly data set where all months consist only four weeks, whenever meet a five-week month, we put off the data of last week to the next month, and regard it as first week in the next month. The original first week then becomes the second. The original second week is then put off to the third, and so on so forth. However, by this way, the first day of some original first weeks may appear in middle of that month. To keep correspondence in fairness in data values, we apply the current monthly data value of macroeconomic indicators when the first day of original first week falls in the 1st to 10th day in the month. When the first day of original first week falls into the 11th to 20th day, we apply the average value of the current and following month. If it appears in any of the last 10 days of a month, we apply monthly indicators' values of next month.

5.2 Measurements

We choose predicting error rate (PER), mean error (ME), root mean square error (RMSE), and mean absolute error (MAE). The definitions of the measures are defined in Table 1.

Table 1

Performance measures and their definitions

Metrics	Calculation
PER	$PER = \left\lvert \frac{T_i - P_i}{T_i} \right\rvert$
ME	$ME = \frac{\sum_{i=1}^{n}(T_i - P_i)}{n}$
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(T_i - P_i)}{n}}$
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert P_i - T_i \rvert$

The T_i and P_i represent the actual and predicted value, respectively, n is total number of samples. All the four measurements are achieved through actual value and forecasting value. The smaller they are, the better the performances are.

5.3 Results

In this section, we show the results of (t-1), (t-2), and (t-3). That is to say, we show the results for the models corresponding to h-step lag (h =1,2,3) for h is the beginning of lags(as explained in 4.1). Then we compare the results of our model with single output approach and multi-output approach with same frequency. For single output approach, the only difference is using basic SVM for four times(in our case) and the other designs are the same as proposed model. For multi-output approach with same frequency, we all adopted weekly data(take time-average of monthly indicator). We use the same RBF as kernel function, and GA as search method in benchmarks. We also use Matlab (2014(a)) and LibSVM (version 1.7) in benchmarks. In our case, $Y_t = [y_{t1}, \dots, y_{tQ}]^T$, $Q = 4$, so $Y_t = [y_{t1}, \dots, y_{t4}]^T$, indicating four weeks in a month. MIMO represents the proposed model, multi-indicator multi output model. MISO represents multi-indicator single output approach with mixed frequency data. SF represents the same frequency approach with multiple outputs. We show the results of last 10 sample points.

Results of (t-1)

Table 2

Per Assessment PER（t-1）

	1st week			2nd week			3rd week			4th week		
	MIMO	MISO	SF	MIMO	MISO	SF	MIMO	MISO	SF	MIMO	MISO	SF
y_1	0.051	0.054	0.053	0.057	0.065	0.056	0.046	0.055	0.053	0.046	0.055	0.055
y_2	0.044	0.048	0.047	0.044	0.052	0.049	0.038	0.043	0.045	0.034	0.044	0.052
y_3	0.034	0.041	0.041	0.033	0.043	0.045	0.030	0.033	0.042	0.032	0.036	0.045
y_4	0.028	0.034	0.036	0.030	0.035	0.040	0.026	0.032	0.039	0.025	0.031	0.039
y_5	0.027	0.031	0.034	0.024	0.026	0.034	0.019	0.027	0.036	0.021	0.029	0.035
y_6	0.024	0.027	0.030	0.019	0.022	0.033	0.017	0.023	0.032	0.018	0.025	0.027
y_7	0.018	0.020	0.026	0.017	0.020	0.027	0.015	0.018	0.028	0.015	0.024	0.025
y_8	0.016	0.017	0.023	0.015	0.017	0.025	0.014	0.018	0.023	0.014	0.019	0.020
y_9	0.013	0.015	0.017	0.014	0.015	0.020	0.012	0.016	0.018	0.013	0.017	0.017
y_{10}	0.011	0.013	0.009	0.011	0.013	0.015	0.011	0.012	0.012	0.010	0.017	0.011
Avg.	0.027	0.030	0.032	0.026	0.031	0.034	0.023	0.028	0.033	0.023	0.030	0.033

Figures 3 and 4 are corresponding to results in Table 2. The full lines represent MISO and SF in Figure 3 and Figure 4, respectively. The dashed lines represent our approach. We compare results among MIMO, MISO and SF. All the approaches show the same trend in results. The proposed approach shows better performances than results of the other two approaches in every week. Compared with MISO, MIMO improves 0.07 at most(the last week), and 0.03 at least(the first week). Compared with SF, MIMO improves 0.1 at most(the last two weeks), and 0.05 at least(the third week).

Figure 3

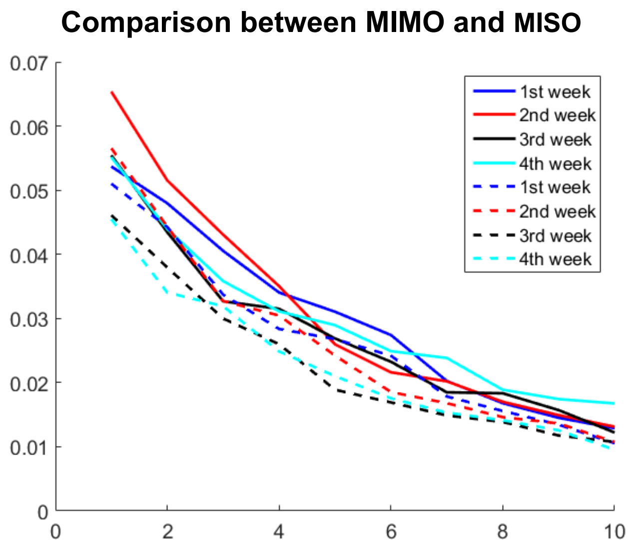


Figure 4

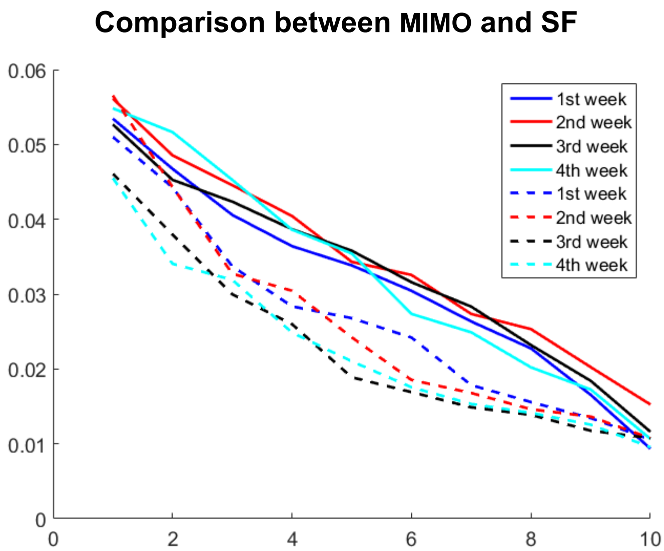


Table 3

Other Results (t-1)

	Model	ME	RMSE	MAE
1st week	MIMO	-9.633	88.016	79.327
	MISO	-8.961	98.04	89.284
	SF	4.512	101.566	94.055
2nd week	MIMO	-27.647	89.551	78.725
	MISO	-41.394	105.187	92.374
	SF	-27.498	109.503	103.194
3rd week	MIMO	-29.735	76.85	68.374
	MISO	10.402	92.245	83.458
	SF	-34.829	104.606	98.023
4th week	MIMO	-16.767	75.129	67.545
	MISO	-31.876	95.682	88.691
	SF	-13.03	106.163	97.005

The ME, RMSE, and MAE all keep in an acceptable level. Though there are some outliers in ME, the whole performances show that MIMO perform better than the other two approaches.

Results of (t-2)

Table 4

Per Assessment PER (t-2)

	1st week			2nd week			3rd week			4th week		
	MIMO	MISO	SF	MIMO	MISO	SF	MIMO	MISO	SF	MIMO	MISO	SF
y_1	0.050	0.051	0.061	0.052	0.054	0.062	0.054	0.058	0.056	0.055	0.056	0.060
y_2	0.044	0.044	0.052	0.041	0.045	0.054	0.045	0.048	0.046	0.043	0.048	0.050
y_3	0.032	0.040	0.036	0.033	0.041	0.035	0.040	0.043	0.041	0.037	0.043	0.041
y_4	0.029	0.030	0.032	0.027	0.031	0.029	0.032	0.032	0.039	0.029	0.032	0.036
y_5	0.017	0.025	0.022	0.018	0.026	0.023	0.026	0.028	0.032	0.024	0.028	0.031
y_6	0.013	0.025	0.017	0.014	0.024	0.021	0.022	0.025	0.026	0.020	0.025	0.025
y_7	0.011	0.019	0.015	0.011	0.020	0.019	0.019	0.019	0.021	0.013	0.021	0.019
y_8	0.010	0.016	0.012	0.011	0.016	0.014	0.015	0.015	0.016	0.012	0.017	0.015
y_9	0.009	0.013	0.013	0.011	0.014	0.012	0.012	0.014	0.016	0.012	0.018	0.015
y_{10}	0.008	0.012	0.009	0.009	0.012	0.008	0.010	0.011	0.014	0.010	0.016	0.010
Avg.	0.022	0.028	0.027	0.023	0.028	0.08	0.028	0.029	0.031	0.026	0.030	0.030

Figure 5

Comparison between MIMO and MISO

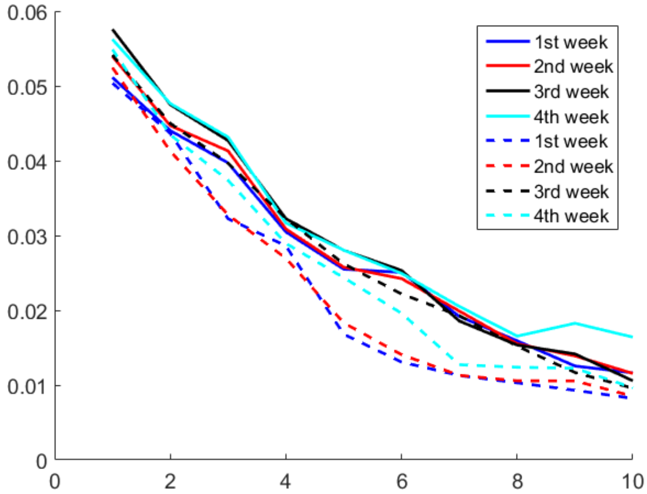
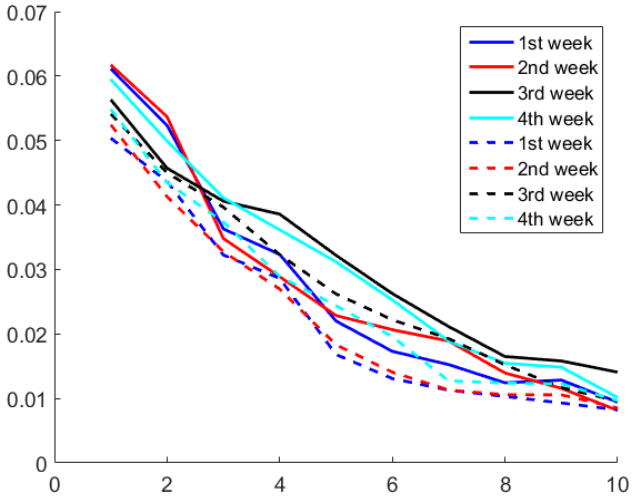


Figure 6

Comparison between MIMO and SF



Figures 5 and 6 are corresponding to results in Table 4. The full lines represent MISO and SF in Figure 5 and Figure 6, respectively. The dashed lines represent our approach. All the approaches show the same trend in results. The proposed approach shows better performances than results of the other two approaches in every week. Compared with MISO, MIMO improves 0.06 at most(the first week), and 0.01 at least(the third week). Compared with SF, MIMO improves 0.05 at most(the first two weeks), and 0.03 at least (the third week).

Table 5

Other Results (t-2)

	Model	ME	RMSE	MAE
1st week	MIMO	-25.524	80.173	66.951
	MISO	25.086	90.876	82.113
	SF	20.44	96.134	81.13
2nd week	MIMO	-0.947	80.791	68.053
	MISO	1.318	93.875	84.469
	SF	38.905	97.089	82.431
3rd week	MIMO	-23.603	93.384	82.63
	MISO	0.143	99	87.725
	SF	16.959	101.194	92.272
4th week	MIMO	-5.956	88.312	76.308
	MISO	35.194	99.546	90.715
	SF	15.743	101.682	90.096

As the same situation in (t-1), there are some outliers in ME, the whole performances show that MIMO perform better than the other two approaches.

Result (t-3)

Table 6

Per Assessment PER (t-3)

	1st week			2nd week			3rd week			4th week		
	MIMO	SIMO	SF	MIMO	MISO	SF	MIMO	MISO	SF	MIMO	MISO	SF
y_1	0.054	0.053	0.056	0.047	0.052	0.052	0.050	0.059	0.064	0.047	0.050	0.059
y_2	0.036	0.040	0.041	0.041	0.043	0.041	0.042	0.053	0.049	0.040	0.045	0.049
y_3	0.031	0.040	0.034	0.036	0.038	0.039	0.038	0.041	0.048	0.038	0.040	0.048
y_4	0.029	0.031	0.033	0.030	0.032	0.034	0.029	0.038	0.044	0.031	0.037	0.046
y_5	0.027	0.031	0.032	0.026	0.030	0.030	0.026	0.034	0.042	0.029	0.030	0.045
y_6	0.023	0.029	0.030	0.022	0.026	0.025	0.019	0.031	0.036	0.024	0.025	0.038
y_7	0.017	0.025	0.025	0.009	0.015	0.016	0.008	0.024	0.031	0.018	0.022	0.032
y_8	0.014	0.017	0.023	0.008	0.011	0.012	0.005	0.014	0.026	0.007	0.018	0.028

	1st week			2nd week			3rd week			4th week		
y_9	0.006	0.011	0.019	0.006	0.008	0.009	0.003	0.011	0.020	0.003	0.012	0.021
y_{10}	0.004	0.005	0.017	0.003	0.005	0.008	0.001	0.005	0.018	0.001	0.006	0.019
Avg.	0.024	0.028	0.031	0.023	0.026	0.027	0.022	0.031	0.038	0.024	0.028	0.038

Figure 7

Comparison between MIMO and MISO

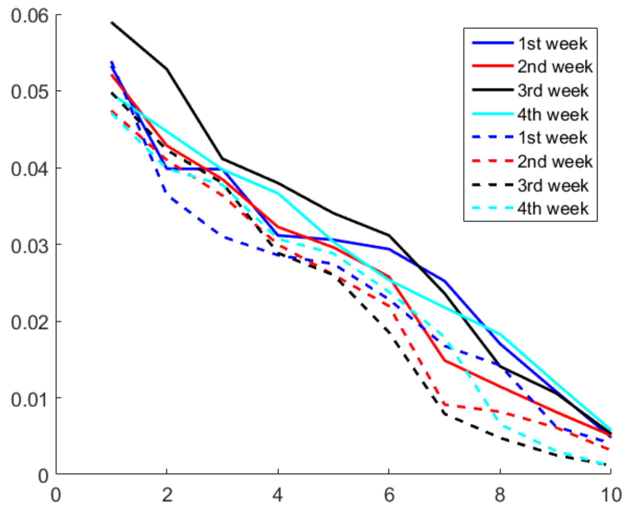
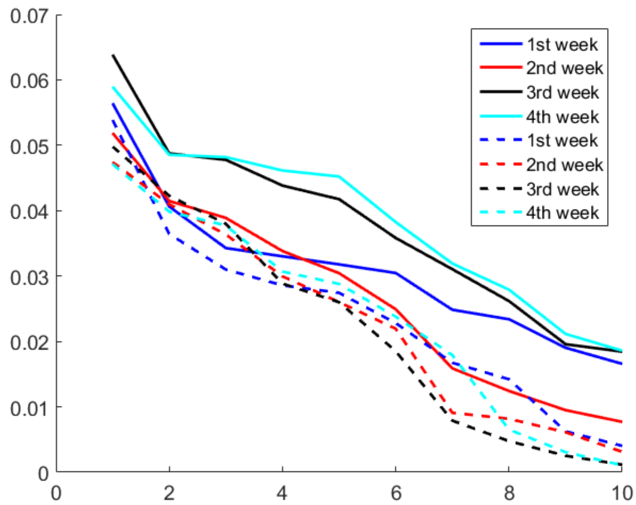


Figure 8

Comparison between MIMO and SF



Figures 7 and 8 are corresponding to results in Table 6. The full lines represent MISO and SF in Figure 7 and Figure 8, respectively. The dashed lines represent our approach. All the approaches show the same trend in results. The proposed approach shows better performances than results of the other two approaches in every week. Compared with MISO, MIMO improves 0.09 at most(the third week), and 0.03 at least(the second week). Compared with SF, MIMO improves 0.16 at most(the third week), and 0.03 at least(the second week).

Table 7

Other Resutls (t-3)

	Model	ME	RMSE	MAE
1st week	MIMO	-3.279	83.569	71.695
	MISO	-2.351	93.343	83.605
	SF	37.741	98.496	92.746
2nd week	MIMO	-33.911	81.982	68.402
	MISO	-0.845	90.076	77.81
	SF	-2.161	90.629	79.811
3rd week	MIMO	-2.326	83.507	92.556
	MISO	-37.722	105.941	83.458
	SF	32.932	120.309	113.058
4th week	MIMO	-33.024	84.125	70.249
	MISO	-12.012	93.881	84.303
	SF	60.501	120.317	114.465

The results are the same as those in (t-1) and (t-2). The MIMO shows better performances overall except outliers in ME.

5.4 DM test

If two models have the same predictability, they don't show significant difference when forecasting. Diebold and Marian(2004) test the above hypothesis by loss function and then compare the performance of models. Suppose $e_{i,t}$ and $e_{j,t}$ represent forecasting errors of model i and j, respectively. $G(\cdot)$ represents the corresponding loss function. The null hypothesis H_0 supposes that if two models have the same predictability, their loss functions have the same expected values. That is, $d_t = g(e_{i,t}) - g(e_{j,t})$, where the expected value $E(d_t) = 0$. $\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$ defines the unbiased estimator of $E(d_t)$. Then

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \sim N(0,1) \tag{16}$$

and $\bar{d} \sim N(0, \hat{V}(\bar{d}))$.

If the value of DM is not significantly different from 0, the two models have the same predictability. If the value of DM is significantly positive, then loss function of model i is significantly positive than that of model j, which means model j have better predictability and vice versa.

Now discuss the statistical significance about the results through DM test. We choose the proposed model as reference model and other models as constrast models. We compare each constrast model with reference model respectively. We algorithm the differences of loss functions between reference model and constrast model. Then we establish statistical tests to test if the difference is significantly different from 0. The loss function here is:

$$Avg.PER = \frac{\sum_{i=1}^{10}\left|\frac{T_i - P_i}{T_i}\right|}{10} \quad (17)$$

Table 8

Results

		Results of MIMO and MISO		Results of MIMO and SF	
		DM	p	DM	p
t-1	1st week	-4.0227	5.75E-05	-5.463	[<0.001]
	2nd week	-2.8871	0.004	-4.8283	[<0.001]
	3rd week	-3.5032	0.00046	-6.0474	[<0.001]
	4th week	-4.6392	3.5E-06	-4.0111	[<0.001]
t-2	1st week	-3.4828	[<0.001]	-2.452*	[0.014]
	2nd week	-5.2488	[<0.001]	-2.3415*	[0.019]
	3rd week	-2.7081	[0.007]	-4.3194	[<0.001]
	4th week	-6.0033	[<0.001]	-4.294	[<0.001]
t-3	1st week	-3.1676	[0.002]	-16.3276	[<0.001]
	2nd week	-3.6903	[<0.001]	-4.1757	[<0.001]
	3rd week	-4.1674	[<0.001]	-7.1028	[<0.001]
	4th week	-4.6209	[<0.001]	-9.3964	[<0.001]

From the p value in second line of every table, we conclude that the results of MIMO and SF at (t-2) have significant difference at 5% level and we mark them “*”. Other results show significant difference at 1% level. The proposed model has better predictability.

6. Conclusion

Researchers usually adopt MIDAS approach to address data sampled at different frequencies. This paper proposes a MIMO-MIDAS forecasting model that involves multi-indicator and gets multiple sequential outputs that have dependencies to each other, and uses small sample size that are more fitful for stock market. The training and evaluation exercises show that the proposed model has better performances and predictability than other models do.

The application of MIMO-MIDAS approach offers a new perspective into use of MIDAS. Though results are satisfactory, it has some limitations. In further researches, we want to figure out whether we can extend to a model including mixed frequencies between indicators. We also try to determine the optimal lag length of indicators. Besides, we want to point out whether it can apply to different economic areas that are inherently full of uncertainties and varieties.

Acknowledgement

The work is supported by the National Natural Science Foundation of China No. 71671019, the Chongqing Basic Research and Frontier Exploration Project of 2019(Chongqing Natural Science Foundation): Grant No., cstc2019jcyj-msxmX0592, and the Science and Technology Research Project of Chongqing Municipal Education Commission in 2019: Grant No, KJQN201900503.

References

Adjaoute, K., Bruand, M. and Gibson R., 1998. On the predictability of the stock market volatility: does history matter?. *European Financial Management,* 4(3), pp.293–319.

Aho, T., Enko, B., Eroski, S. and Elomaa, T., 2012. Multi-target regression with rule ensembles. *Journal of Machine Learning Research*,13(1), pp.2367-2407.

Andreou, E. *et al.*, 2019. Inference in Group Factor Models With an Application to Mixed frequency Data[J]. *Econometrica,* 87(4), pp. 1267-1305.

Asgharian, H., Hou, A.J. and Javed F., 2013. The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH‐MIDAS Approach. *Journal of Forecasting,* 32(7), pp.600-612.

Bao, Y. Xiong, T. and Hu, Z., 2014. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, 129, pp.482–493.

Bessec, M., 2017. Revisiting the transitional dynamics of business-cycle phases with mixed-frequency data. *Econometric Reviews*, 38(7), pp.711-732.

Blockeel, H. and Raedt, L.D., 1998. Top-down induction of first-order logical decision trees. *Artificial Intelligence,* 101(1–2), pp. 285-297.

Breiman, L. and Friedman, J.H., 1997. Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society,* 59(1), pp. 3-54.

Caruana, R., 1997. Multitask Learning. *Machine Learning*, 28(1), pp. 41-75.

Chang, T.Y. Fang, H. and Lee, Y.H., 2015. Nonlinear A Djustment To The Long-Run Equilibrium Between The Reit And The Stock Markets In Japan And Singapore. *Romanian Journal for Economic Forecasting*, 18(3), pp.27-38.

Chen, X. and Ghysels, E., 2011. News—good or bad—and its impact on volatility predictions over multiple horizons. *Review of Financial Studies*, 24, pp. 46-81.

Cheng, S.C. and Wu, T.P., 2013. Nonlinear Behavior of the US Stock Price-Dividend: Evidence from Threshold Unit Root Tests. *Romanian Journal of Economic Forecasting*, 16(4), pp.82-93.

Crammer, K. and Singer, Y., 2002. On the algorithmic implementation of multiclass kernel-based vector machines. *The Journal of Machine Learning Research*, 2, pp.265-292.

Duarte, C., Rodrigues, P.M.M. and Rua, A., 2017. A mixed frequency approach to the forecasting of private consumption with atm/pos data. *International Journal of Forecasting*, 33(1), pp.61-75.

Diebold, F.X. and Mariano, R.S., 2002. Comparing predictive accuracy.*Journal of Business & Economic Statistics, 20*(1), pp.134-144.

Han, Z., Ying Liu; Jun Zhao and Wei Wang., 2012a. Real Time Prediction for Converter Gas Tank Levels Based on Multi-Output Least Square Support Vector Regressor. *Control Engineering Practice*, 20(12), 1400-09.

____ 2012b. Real Time Prediction for Converter Gas Tank Levels Based on Multi-Output Least Square Support Vector Regressor. *Control Engineering Practice*, 20(12), pp.1400–09.

Hepenstrick, C. and Marcellino, M., 2018. Forecasting gross domestic product growth with large unbalanced data sets: the mixed frequency three-pass regression filter. *Journal of the Royal Statistical Society: Series A (Statistics in Society).*

Foroni, C., Marcellino, M. and Schumacher, C., 2015. Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 178, pp.57-82.

Forsberg, L. and Ghysels, E., 2007. Why do absolute returns predict volatility so well?. *Journal of Financial Econometrics*, 5, pp. 31-67.

Galvão, A.B., 2013. Changes in predictive ability with mixed frequency data. *International Journal of Forecasting,* 29, pp. 395-410.

Ghysels, E. and Ozkan, N., 2015. Real-time forecasting of the us federal government budget: a simple mixed frequency data regression approach. *International Journal of Forecasting*, 31(4), 1009-1020.

Ghysels, E. Santa-Clara, P. and Valkanov, R. 2004. The MIDAS touch: Mixed data sampling regression models. *Cirano Working Papers,* 5(1), pp. 512-517.

Ghysels, E. Santa-Clara, P. and Valkanov, R., 2005. There is a risk-return trade-off after all. *Journal of Financial Economics*, 76, pp. 509-548.

Ghysels, E., Santa-Clara, P. and Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1), pp. 59-95.

Han, Z., Liu, Y. and Zhao, J., 2012. Real time prediction for converter gas tank levels based on multi-output least square support vector regressor. *Control Engineering Practice*, 20(12), pp. 1400-1409.

Hiemstra, C. and Jones, J.D., 1994. Testing for linear and nonlinear Granger causality in the stock price ‐ volume relation. The Journal of Finance, 49, pp. 1639-1664.

Kocev, D. *et al*., 2009. Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. *Ecological Modelling,* 220(8), pp. 1159-1168.

León, A. Nave, J. M. and Rubio, G., 2007. The relationship between risk and expected return in Europe. *Journal of Banking & Finance, 31*(2), pp. 495-512.

Lupu, I. 2015. European Stock Markets Correlations In A Markov Switching Framework. *Romanian Journal of Economic Forecasting*, 18(3), pp. 103-119.

Mao, W. *et al*., 2014a. Leave-One-out Cross-Validation-Based Model Selection for Multi-Input Multi-Output Support Vector Machine. *Neural Computing and Applications*, 24(2), 441-51.

Mao, W., Xu, J. and Wang, C., 2014b. A fast and robust model selection algorithm for multi-input multi-output support vector machine. *Neurocomputing*, 130, pp 10-19.

Nedaie, A. and Najafi, A.A., 2016. Polar support vector machine: Single and multiple outputs. *Neurocomputing*, 171, pp. 118-126.

Pérez-Cruz, F., Camps-Valls, G. and Soria-Olivas, E., 2002. Multi-dimensional function approximation and regression estimation. *Artificial Neural Networks—ICANN*, Springer Berlin Heidelberg, pp. 757-762.

Perez-Cruz, F. and Artes-Rodriguez, A., 2001. A new optimizing procedure for nu-support vector regressor. *In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, (ICASSP '01), IEEE, Salt Lake City, UT, pp. 1265–1268.

Platt, J.C., Cristianini, N. and Shawe-Taylor, J., 1999. Large margin dags for multiclass classification. *Advances in Neural Information Processing Systems, 12*(3), pp. 547--553.

Sánchez-Fernández, M. *et al*., 2004. SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. *Signal Processing, IEEE Transactions on,* 52, pp. 2298-2307.

Schölkopf, B. and Smola, A.J., 2002. Learning with kernels: Support vector machines, regularization, optimization, and beyond, *MIT press*.

Suykens, J.A.K. and Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural processing letters*, 9(3), pp. 293-300.

Suzuki, E., Gotoh, M. and Choki, Y., 2001. Bloomy Decision Tree for Multi-objective Classification. *European Conference on Principles of Data Mining and Knowledge Discovery,* Springer-Verlag, 2168, pp. 436-447.

Thenmozhi, M., Chand, G.S., 2016. Forecasting stock returns based on information transmission across global markets using support vector machines. *Neural Computing and Applications*, 27(4), pp. 805-824.

Tsui, A.K., Xu, C.Y. and Zhang, Z., 2018. Macroeconomic forecasting with mixed data sampling frequencies: evidence from a small open economy. *Journal of Forecasting*.

Tuia, D. *et al*., 2011. Multioutput support vector regression for remote sensing biophysical parameter estimation. *Geoscience and Remote Sensing Letters, IEEE,* 8, pp. 804-808.

Vapnik, V., 1995. The nature of statistical learning theory. *Springer Science & Business Media.*

Vapnik,V., 1998. *Statistical Learning Theory*. New York: Wiley.

Wang, K., 2014. Modeling Stock Index Returns using Semi-Parametric Approach with Multiplicative Adjustment. *Romanian Journal of Economic Forecasting*, 17(4), pp. 65-75.

Wang, X. and Xiao, Z., 2017. Regional eco-efficiency prediction with support vector spatial dynamic midas. *Journal of Cleaner Production*. 161, pp.165-177.

Weston, J. and Watkins, C. 1999. Support vector machines for multi-class pattern recognition. *//ESANN*. 99, pp. 219-224.

Xiong, T., Bao, Y. and Hu, Z., 2014. Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting. *Knowledge-Based Systems,* 55, pp. 87-100.

Xu, Q., Zhuo, X., Jiang, C. and Liu, Y., 2018. An artificial neural network for mixed frequency data. *Expert Systems with Applications*.

Żbikowski, K., 2015. Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications*, 42(4), pp.1797-1805..