Expert Systems with Applications 38 (2011) 1697–1701

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Genetic algorithms for the investment of the mutual fund with global trend indicator

Tsung-Jung Tsai, Chang-Biau Yang *, Yung-Hsing Peng

Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan

ARTICLE INFO

Keywords:
Fund
Indicator
Genetic
Algorithm
Investment strategy

ABSTRACT

Our investment strategy for the world mutual funds can be divided into three main parts. First, the global trend indicator (GTI) is defined for evaluating the price change trend of the funds in the future. Then, based on GTI, we derive the monitoring indicator (MI) to measure whether the fund market is in the bull or bear state. Finally, to decide the signal for buying or selling funds, a genetic algorithm is invoked to dynamically select funds according to their past performances (profitability). In our experimental results from January 1999 to December 2008 (10 years in total), we achieve the annual profit higher than 10%.

1. Introduction

Mutual fund is a very popular investment tool. A fund collects the capital of the investors, and then its manager use the capital to invest in the stock markets, foreign exchange markets, bond markets and so on. However, the financial market is chaotic, complex and full of uncertainty. Hence, making money in the financial market is a very difficult task that is worth study.

To make money in the financial market, some of researchers usually analyze these time series with some useful tools such as *support vector machine* (SVM) (Huang, Nakamori, & Wang, 2005; Ince & Trafalis, 2008; Pai & Lin, 2005; Tay & Cao, 2001), *genetic complementary learning* (GCL) (Tan, Quek, & Ng, 2005), *genetic algorithms* (GA) (Kim & Han, 2000) and *neural networks* (NN) (Kim & Han, 2000). In these ways, they try to predict the price or trend in the future. The goals of these researchers are to raise the prediction accuracy and to make more money in the stock market, based on their learning models.

From the aspect of measuring stocks, some researchers use *root mean square deviation* (RMSE) (Ince & Trafalis, 2008; Pai & Lin, 2005) and *normalized mean squared error* (NMSE) (Tay & Cao, 2001) to measure the performance of their stock forecast. Others adopted the prediction accuracy (Huang et al., 2005; Kim & Han, 2000; Tan et al., 2005) to evaluate their experimental results. However, we want to address that most researchers do not consider the transaction fee, which has a strong impact on the profit of the investment. Therefore, one can see that the these models cannot be applied to general stock markets. In this paper, we propose a new investment strategy that achieves high annual return when considering the transaction fee. Hence, compared with other existing results, our strategy is more suitable for real investment.

The rest of this paper is organized as follows. Section 2 explains required notations and techniques. Next, we propose our main algorithm in Section 3 and show the experimental results in Section 4. Finally, we give our conclusions in Section 5.

2. Preliminary

2.1. The ROI and average annualized return

We use the notation $ROI_{t_1}^{t_2}(x)$ to represent the performance of a fund x from day t_1 to t_2. For easy description, we adopt $ROI_T(x,t)$ to represent the performance of x for the latest period T dating back from current day t. In our experiments, we refer to the *average annualized return* as a fair comparison for various strategies. The formula of the average annualized return is described as follows:

$$FV = PV(1+\bar{r})^n,$$

where PV (present value) is the initial capital, FV (future value) is the return capital, $\bar{r}$ is the average annualized return, and n is the number of years. The average annualized return is also known as the *geometric average return*.

2.2. Technical analysis – SMA (simple moving average)

Simple moving average (SMA) (StockCharts.com, 1999) is a very popular and easy-to-use tool. SMA computes the average of recent prices over a period. SMA can be used to smoothen the series of recent prices, and to reduce the volatility of prices. The formula of SMA is given as follows:

* Corresponding author.
E-mail address: cbyang@cse.nsysu.edu.tw (C.-B. Yang).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.07.094

$$SMA_T(t) = \sum_{i=0}^{T-1} price(t-i) \Big/ T,$$

where T is a period for calculating prices, t represents the current day, and $price(t)$ represents the price at day t. In Section 3, we will use SMA to design our indicators.

2.3. Genetic algorithms

The *genetic algorithm* (GA) (Someya & Yamamura, 2001) is a tool for handling optimization problems (Holland, 1975). GA is one kind of evolutionary algorithms, which simulates the survival principle (survival of the fittest) of creatures in the nature. In GA, the natural environment is simulated by computational operations including *selection*, *crossover* and *mutation*. In the next section, we will adopt GA to select good funds for investment.

3. Our fund investment strategy

In a *bear market*, buying mutual funds is forbidden and all of the investor's fund holdings are redeemed. Our model allows funds to be invested in a *bull market*. In the buying procedure, the priorities of funds are determined by genetic algorithms. We rank all funds by the ordered priorities and create a ranking list. With the ranking list, we redeem funds if their ranks do not locate within the *replacement threshold* γ, which is predefined by the user. Here, with the replacement scheme, we can always keep better funds and remove bad funds. Then new funds will be picked sequentially (according to their ranks) until the size of our portfolio reaches the predefined limit N. In our portfolio, we repeat this procedure until the end of investment.

In the following subsections, we first define the*global trend indicator* (GTI) and the *monitoring indicator* (MI). And then, we explain how to select funds with genetic algorithms.

3.1. The global trend indicator

Let F denote the set of funds to be considered. The formula of our *global trend indicator* (GTI) is given as follows:

$$\begin{aligned}
&ROI_{1D}(x,t) = \frac{NAV_x(t)}{NAV_x(t-1)} - 1, \quad x \in F,\\
&U(t) = \{x | NAV_x(t) > NAV_x(t-1),\ x \in F\},\\
&D(t) = \{x | NAV_x(t) \leqslant NAV_x(t-1),\ x \in F\},\\
&UM(t) = \frac{\sum_{x \in U(t)} ROI_{1D}(x,t)}{\sum_{x \in F} |ROI_{1D}(x,t)|},\\
&DM(t) = \frac{\sum_{x \in D(t)} ROI_{1D}(x,t)}{\sum_{x \in F} |ROI_{1D}(x,t)|},\\
&GTI(t) = GTI(t-1) \times \{1 + UM(t) + DM(t)\},\\
&GTI(0) = 100 \ (\text{base value}),
\end{aligned}$$

where t represents a trading day, $NAV_x(t)$ represents the NAV for fund x at day t, $ROI_{1D}(x,t)$ (the *daily return*) means the percentage of the price change for the fund x from day $t-1$ to day t.

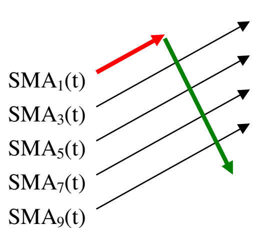

Fig. 1. The reverse pattern from the bull market to the bear market.

3.2. The monitoring indicator

The *monitoring indicator* (MI) is derived from GTI and its four simple moving average indicators. The period of the four moving averages are 3, 5, 7 and 9, denoted as $SMA_3(t)$, $SMA_5(t)$, $SMA_7(t)$ and $SMA_9(t)$, respectively. The formula of our monitoring indicator $MI(t)$ is given as follows:

$$\begin{aligned}
&MI_{win}(t) = \sum_{1 \leqslant i < j \leqslant 5} \{j - i | SMA_{2i-1}(t) > SMA_{2j-1}(t)\}/20,\\
&MI_{lose}(t) = -\sum_{1 \leqslant i < j \leqslant 5} \{j - i | SMA_{2i-1}(t) \leqslant SMA_{2j-1}(t)\}/20,\\
&MI(t) = MI_{win}(t) + MI_{lose}(t).
\end{aligned}$$

Next, we describe how to use MI to derive the selling-and-buying signals. For the selling signals, we sample the values of weekly GTI. The sampled values change a lot when the financial situation reverses from the bull market (a bull ranking in parallel) to the bear market, as depicted in Fig. 1. We set the conditions to identify the selling pattern as follows:

$$\begin{aligned}
&triggerSignal_{1W}(t)\\
&= \begin{cases} sell\ if\ ((MI_{1W}(t-14) = 1\ or\ MI_{1W}(t-7) = 1)\ and \\ \qquad (SMA_1(t) < SMA_9(t)\ or\ SMA_1(t-7) < SMA_9(t-7))), \\ unknown\ otherwise, \end{cases}
\end{aligned}$$

$$\begin{aligned}
&triggerSignal_{2W}(t)\\
&= \begin{cases} sell\ if\ ((MI_{2W}(t-28) = 1\ or\ MI_{2W}(t-14) = 1)\ and \\ \qquad (SMA_1(t) < SMA_9(t)\ or\ SMA_1(t-14) < SMA_9(t-14))), \\ unknown\ otherwise. \end{cases}
\end{aligned}$$

For determining the buying signals, we sample the values of monthly GTI and use them to establish the one-month monitoring indicator MI_{1M}

$$triggerSignal_{1M}(t) = \begin{cases} buy & \text{if } MI_{1M}(t) = 1, \\ unknown & \text{otherwise}. \end{cases}$$

Let $signal(t)$ be the buy-and-sell signal at day t. With the combination of $triggerSignal_{1W}(t)$, $triggerSignal_{2W}(t)$, and $triggerSignal_{1M}(t)$, we can determine $signal(t)$.

The default transaction fee is set to 3%, which can be redefined by users. In a bull market, $signal(t) = sell$ indicates that the bear market is coming. For this case, our model will redeem all fund holdings (take back all money). Later on, any buying operation will be forbidden until $signal(t) = buy$. On the contrary, if $signal(t)$ turns to *buy* in a bear market, it indicates that the bull market is coming. At this time, our model will start to buy good funds. After that, the buying operation is allowed until $signal(t)$ turns to *sell*.

3.3. Fund selection with genetic algorithms

When a buying signal appears, we have to pick out good funds from F as our investment objects and put them into the portfolio list with size N. That is to say, exactly N funds are invested. To estimate the potential profitability for funds, we define a scoring function that considers the performance indicators of all funds, during the past 3, 6, 9 months, and 1, 2, 3, 4, 5 years. The formula of the scoring function *score* is given as follows:

$$\begin{aligned}
&score(x,t)\\
&= \frac{w_{3M}(t)}{rank(ROI_{3M}(x,t))} + \frac{w_{6M}(t)}{rank(ROI_{6M}(x,t))} + \frac{w_{9M}(t)}{rank(ROI_{9M}(x,t))}\\
&\quad + \frac{w_{1Y}(t)}{rank(ROI_{1Y}(x,t))} + \frac{w_{2Y}(t)}{rank(ROI_{2Y}(x,t))} + \frac{w_{3Y}(t)}{rank(ROI_{3Y}(x,t))}\\
&\quad + \frac{w_{4Y}(t)}{rank(ROI_{4Y}(x,t))} + \frac{w_{5Y}(t)}{rank(ROI_{5Y}(x,t))},
\end{aligned}$$

where $w_{3M}(t), w_{6M}(t), \ldots, w_{5Y}(t)$ represent the weights of each ROI of fund x, which are limited in the range $[0,1]$. In addition, $rank(ROI_T(x,t))$ means the rank of ROI_T of fund x on day t among all considered funds. According to this scoring function, the rank of each fund varies with different weights. Therefore, to obtain suitable weights for ranking funds, a genetic algorithm is invoked.

In each generation of our genetic algorithm, there are several chromosomes in the population. Each chromosome can be decoded into eight weights for performance indicators $ROI_{3M}(x,t)$, $ROI_{6M}(x,t)$, $ROI_{9M}(x,t)$, $ROI_{1Y}(x,t)$, $ROI_{2Y}(x,t)$, $ROI_{3Y}(x,t)$, $ROI_{4Y}(x,t)$, and $ROI_{5Y}(x,t)$. We use these weights to calculate the score of each fund by the scoring function. We put the first N highest funds into the portfolio.

If the poor funds have not yet been weeded out, the size of the portfolio would remain unchanged, and there is no place to put new funds. Here we define a *replacement threshold*, denoted as γ, for weeding out poor funds. Taking $|F| = 50$ and $\gamma = 40\%$ for example, once the rank of a certain fund falls outside top 20 (=50 $\times$ 40%), this fund will be redeemed and replaced with a better one having higher priority.

We repeat the training until day t_{end} (the end of the training). The evolutionary process will try to find the best profit and report the best weights to the system. After training, for a short period of time we can use these best weights to select and invest good funds.

4. Experimental results

In this section, we will compare our strategy with other well-known strategies. Our test data set is taken from Fund DJ (http://www.funddj.com/). We select only equity funds with A-type and US dollar. We assume that all of the compared strategies have the same capital in the initial state. In our experiment, we set the initial capital to $1 on 1999/1/1, and evaluate the ending-period capital at the end of 2008. In the following, we use the term ROI_{cum} to denote the cumulative ROI from 1999/1/1 to 2008/12/31.

Table 1
The symbol table for various constant weights.

	w_{3M}	w_{6M}	w_{9M}	w_{1Y}	w_{2Y}	w_{3Y}	w_{4Y}	w_{5Y}
–	1	1	1	1	1	1	1	1
\	1	0.857	0.714	0.571	0.429	0.286	0.143	0
/	0	0.143	0.286	0.429	0.571	0.714	0.857	1
∧	0	0.333	0.667	1	1	0.667	0.333	0
V	1	0.667	0.333	0	0	0.333	0.667	1
W	1	0	0	1	1	0	0	1
M	0	1	1	0	0	1	1	0

4.1. The motley fool investment

The motley fool investment is a very simple and lazy investment rule. At the beginning of every year, we select the only fund whose one-year *ROI* (the performance of the previous year) is the best, and put it into the portfolio. At the end of a year, we redeem all fund holdings. The redeemable capital of the funds will be taken as the initial capital for the next year.

4.2. The 4433 rule

The *4433 rule* (FundDJ Co., Ltd., 2000) for investing fund is a simple investment rule, which is well-known in Taiwan. Based on the rule, we can obtain the ending-period capital $2.0768 from 1999/1/1 to 2008/12/31.

4.3. The best buy-and-hold strategy for ten years

In this strategy, it is assumed that there exists a very smart investor, who knows which fund has the best performance for ten years. The investment starts from 1999/1/1, and the fund is held until 2008/12/31. If the launch day of this fund is after 1999/1/1, then he would postpone his investment on its launch day.

4.4. The performance of the one-year time deposit

We assume the interest rate of one-year time deposit is 3%. The performance calculation is given as follows:

$$\text{Future Value} = \$1 \times (1+3\%)^{10} = \$1.3439, \quad ROI_{cum} = 34.39\%.$$

4.5. The MSCI world price index and theS&P 500 composite price index

We take the MSCI (*Morgan Stanley Capital International*) world price index (Taylor, 1990) as one of our contrasts. The close prices on 1999/1/1 and 2008/12/31 are 1,149.95 and 920.2, respectively.

As another contrast, we refer to the S&P 500 (*Standard and Poor's 500*) composite price index (Taylor, 1990). The close price on 1999/1/4 and 2008/12/31 are 1,228.10 and 903.25, respectively.

4.6. Our strategy with user-defined weights

In these experimental results, we only refer to the MI (monitoring indicator) derived from our GTI (global trend indicator) to buy and sell funds, but do not import the genetic algorithm to our mod-

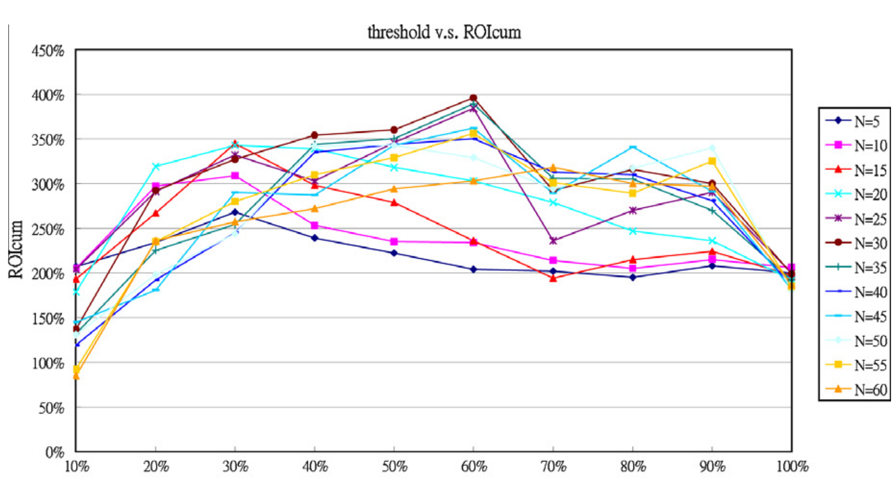

Fig. 2. The relationship between γ and 10-year cumulative ROI's with user-defined constant weights for various N.

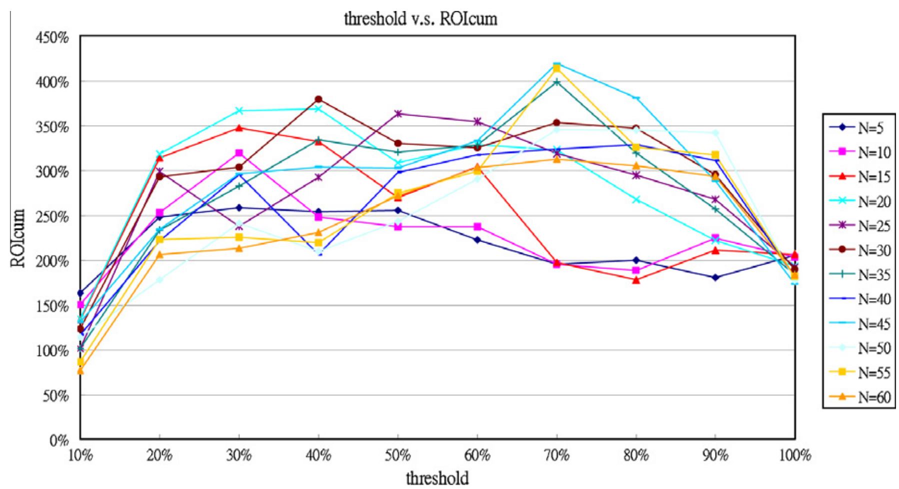

Fig. 3. The relationship between γ and 10-year cumulative ROI's based on GA with various N.

el. We define various constant weights in Table 1, and obtain the relationship between the replacement threshold and cumulative ROI shown in Fig. 2. One can see that 10-year cumulative ROI's are located between 150% and 400% if $20\% \leqslant \gamma \leqslant 90\%$.

Table 2
10-year cumulative ROI's of the user-defined constant weights, the dynamic weights obtained by GA, and the constant weights obtained by GA.

N:	45	55	35
γ:	70%	70%	70%
-	**357%**	**361%**	304%
\	240%	296%	197%
/	252%	279%	150%
∧	267%	314%	**375%**
V	175%	221%	179%
w	**449%**	289%	**663%**
M	281%	**344%**	274%
Average ROI's for user-defined constant weights	289%	301%	306%
Average ROI's for dynamic weights based on GA	**420%**	**414%**	**399%**
ROI's for constant weights based on GA	314%	312%	373%

Table 3
Performance comparison for various strategies, where ROI_{cum} and AR_{cum} denote the cumulative and annualized ROI from 1999/1/1 to 2008/12/31, respectively.

Strategy	ROI_{cum} (%)	AR_{cum} (%)
The motley fool investment	−8.52	−0.89
The 4433 rule	127.69	8.58
The buy-and-hold strategy with the maximum ROI_{cum}	562.19	20.81
The buy-and-hold strategy with the 10th maximum ROI_{cum}	256.19	13.55
One-year time deposit	34.39	3.00
MSCI world price index	−19.98	−2.20
S&P 500 composite price index	−26.45	−3.03
User-defined weights – best ROI_{cum} ($N = 30, \gamma = 60\%$)	800	24.57
User-defined weights – best average ROI_{cum} ($N = 30, \gamma = 60\%$)	396	17.37
User-defined weights – worst average ROI_{cum} ($N = 45, \gamma = 20\%$)	181	10.88
User-defined weights – worst ROI_{cum} ($N = 5, \gamma = 20\%$)	85	6.35
Constant weights (GA) ($N = 45, \gamma = 70\%$)	313	15.24
Constant weights (GA) ($N = 55, \gamma = 70\%$)	312	15.21
Constant weights (GA) ($N = 35, \gamma = 70\%$)	373	16.81
Dynamic weights (GA) – best ROI_{cum} ($N = 20, \gamma = 70\%$)	954	26.56
Dynamic weights (GA) – best average ROI_{cum} ($N = 45, \gamma = 70\%$)	420	17.92
Dynamic weights (GA) – worst average ROI_{cum} ($N = 50, \gamma = 20\%$)	178	10.77
Dynamic weights (GA) – worst ROI_{cum} ($N = 50, \gamma = 30\%$)	66	5.2

4.7. Our strategy with dynamic weights obtained by GA

In these experiments, we try to obtain superior weights with genetic algorithms. In our GA, each chromosome is formed by 8 weights, where each weight is encoded by 6 bits. The number of generations is set to 25 and the population size is set to 80. Each GA repeats three times to obtain three kinds of superior weights. Finally, we take the best weight that has maximum fitness (ROI). In this way, we avoid the local optima. With various replacement thresholds and the portfolio sizes N, we organize the average cumulative ROI in Fig. 3, showing the relationship between the replacement threshold and cumulative ROI. One can see that 10-year cumulative ROI's based on genetic algorithms are located between 150% and 450% if $20\% \leqslant \gamma \leqslant 90\%$.

4.8. Our strategy with constant weights obtained by GA

According to the experimental results of Section 4.7, we find that the best parameters of the best three ROI_{cum} (420%, 414% and 399%) are ($N = 45, \gamma = 70\%$), ($N = 55, \gamma = 70\%$) and ($N = 35, \gamma = 70\%$), respectively.

With these best parameters, we try to find out the best constant weights, which can be used throughout the whole investment. Therefore, the genetic algorithm is only applied to find the weights at the beginning of the investment. In other words, the weights are not changed in the subsequent selling-and-buying operations. By doing so, we avoid frequent training. Here we use the trading data in the latest three years as our training data for this GA.

Table 2 shows the comparison of the performance between the dynamic weights obtained by GA and the constant weights obtained by GA. The value of each entry represents a 10-year cumulative ROI from 1999 to 2008. One can see that dynamic weights achieve higher profit than constant weights obtained by GA. However, the constant weights obtained by GA usually outperform the user-defined constant weights.

In Table 3, we summarize the performance of various strategies. For strategies using dynamic weights, we do 10 individual experiments to obtain their best, averaged, and worst ROI's.

5. Conclusion

In this paper, we discuss how to select funds with potential profitability, and to indicate the time points for buying or selling funds. In this way, we can avoid the financial turmoil or turbulence, and reduce the investment loss. Our experiments are different from those done by other researchers, since we take transaction fees into account. We suggest investors to take the investment strategy with dynamic weights based on GA, which makes an excellent profit (high annual return) in our simulation. One can see that our strategy outperforms others, because it provides good buying-and-selling points to investors.

References

FundDJ Co., Ltd. (2000). FundDJ. Available from http://www.funddj.com/.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Michigan: University of Michigan Press.

Huang, W., Nakamori, Y., & Wang, S.-Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research, 32*(10), 2513–2522.

Ince, H., & Trafalis, T. B. (2008). Short term forecasting with support vector machines and application to stock price prediction. *International Journal of General Systems, 37*(6), 677–687.

Kim, K.-J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications, 19*, 125–132.

Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *OMEGA: The International Journal of Management Science, 33*, 497–505.

Someya, H., & Yamamura, M. (2001). Genetic algorithm with search area adaptation for the function optimization and its experimental analysis. In *Proceedings of the 2001 congress on evolutionary computation, Seoul, South Korea* (Vol. 2, pp. 933–940).

StockCharts.com. (1999). ChartSchool. Available from http://stockcharts.com/.

Tan, T. Z., Quek, C., & Ng, G. S. (2005). Brain-inspired genetic complementary learning for stock market prediction. In *Proceedings of IEEE congress of evolutionary computing, Edinburgh, UK* (Vol. 3, pp. 2653–2660).

Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *OMEGA: The International Journal of Management Science, 29*(4), 309–317.

Taylor, B. (1990). GlobalFinancialData. Available from http://www.globalfinancialdata.com.