MODELING STOCK RETURN IN UNITED STATES AND CHINA BY USING GLS METHOD AND AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY MODEL

A THESIS

Presented to

The Faculty of the Department of Economics and Business

The Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

By

Junhao(Billy) Mai

March 2022

MODELING STOCK RETURN IN UNITED STATES AND CHINA BY USING GLS METHOD AND AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY MODEL

Junhao (Billy) Mai

March 2022

Mathematical Economics

Abstract

This paper tries to model the stock return in the United States and China separately, using SP 500 index and CSI 300 index respectively as modeling subjects, by adopting the Generalized Least Square (GLS) model and Autoregressive Conditional Heteroskedasticity (ARCH) model. The closing yield of 10Y treasury bond and 10Y corporate bond, the closing price of commodity index, and interbank offer rate overnight are employed as independent variables. Residuals from two models are formulated by the ARCH model. This process visualizes and predicts the volatility in the market. The models reveal the relationship between the performance of other asset classes and stock return in the United States and China. These will help investors and portfolio managers to allocate their positions and guide their investment strategy by forecasting the volatility in the market.

KEYWORDS: (Modeling, Stock Return, Volatility, Treasury yield, Commodity price)
JEL CODES: (G10, G11, G12)

Acknowledgement

I would like to thank my thesis advisor Professor Oguzhan Batmaz for his help during my thesis blocks. I also want to thank my Statistics Professor Flavia Sancier-Barbosa, Econometric Professor Aju Fenn, Ordinary Differential Equation Professor Andrea Bruder for their intriguing and amazing classes in Mathematics and Economics. Finally, I would like to thank my friends, my family, and everyone who helped and supported me in my four-year college life.

Hope this thesis will be helpful to anyone who reads it !

ON MY HONOR, I HAVE NEITHER GIVEN NOR RECEIVED UNAUTHORIZED AID ON THIS THESIS

Zunhao Mai

Signature

TABLE OF CONTENTS

Introduction

Riding the crest of financial development and the booming of asset management, the performance of stock market has been monitored by many investors, ranging from institutional portfolio managers to individual traders. Investing in stock is also considered as one of the most efficient strategies of asset management and value creation. According to the data from World Bank (2021), the global stock capitalization has been nearly doubled in the past ten years, increased from 47.52 trillion USD in 2011 to 93.69 trillion USD in 2020. By the end of 2020 (2021), the total market capitalization accounts for 134.65% of the world GDP.

Therefore, modeling stock return and try to beat the market have been the dream for investors around the world. Many models with various different parameters have been created in the past. This study tries to model the stock market return by using two other common asset classes: treasury and commodity. Report from Treasury market is even bigger than stock market with global market capitalization of 123.5 trillion USD in 2021 (Kolchin et al., 2021). Treasury yield in the United States is widely considered as the risk-free rate for financial instruments with distinct time maturity worldwide. For example, the US 10Y treasury bond yield is represented as the long-run risk-free rate. In various stock valuation methods, like Discounted Cash Flow Valuation Method, they all use market-determined rate to price stocks. Thus, this indicator may have a huge effect on the stock market return.

Additionally, commodity price is a crucial asset class in the industry as well. Ranging from industrial raw material, such as copper and aluminum, to consumer supplies raw materials, like cotton and egg. Commodities are closely related to the real economy and the price fluctuation will definitely affect the performance of the stock market which is believed as the barometer of the real economy. For example, during year 2021, the effect of quantitative easing has fueled the prosperity of commodity market as well as the equity market. These two asset classes demonstrate strong correlation. In addition, two important variables, overnight interbank offered rate and corporate bond yield will also be added to the model to represent the significance of liquidity and credit variation on stock return.

This study will model the stock return in Chinese and US stock market in the most recent ten years (2012-2021). China and United States are the two largest economies in the world and the total capitalization of both countries' stock market exceeds 50 trillion USD, which accounts for more than 50% of the global stock market capitalization. Selecting these two countries' stock index as sample space will be highly representative. What's more, this study chooses 2012-2021 as the time span to model the stock return in order to create a most up-to-date model.

One of the problems that confuses and attracts many investors is to model the stock return and try to gain excess yield. This study will try to build a model for stock return by using treasury yield and commodity prices along with other significant macro variables. Apart from traditional studies which only focus on the relationship between stock return and treasury yield or commodities, this study tries to use treasury yield and commodity together to better model the stock return. Different from other studies, the relationship between the stock market return and treasury yields as well as commodity price in China and the United States will be closely examined. Generalized Least Square Method (GLS) will be employed to estimate the preliminary models and the residuals from the model will be examined by Autoregressive conditional heteroscedastic method (ARCH) in order to construct the model for stock return and the level of risk embedded in the market. This may be helpful for portfolio managers and individual investors to monitor the risk of the market and the correlation of different asset classes. Based on common logic, stock performance should be highly related to treasury yields and commodity price in both countries, with positive correlation to both treasury yields and commodity price in both countries' economy.

Literature Review

Variable choosing has been a vital process for modeling stock returns in different literature papers in the past. Most of the paper examines the predictability of macro indicators such as dividend yields (Ang and Bekaert, 2007), short-term and long-term interest rates as well as the consumption-wealth ratio (Campbell and Thompson, 2007). In 2007, Ang and Bekaert investigated the significance of dividend yields to predict the excess return. They concluded that dividend yields have predictive power in short run, but have no predictive power in long run. Campbell and Thompson later discovered that many variables, such as "stock market valuation ratios, short- and long-term interest rates, patterns in corporate finance or the cross-sectional pricing of individual stocks, or the level of consumption in relation to wealth", have better predictive power than "historical average return forecast" in out-of-sample test.

Besides these regular indicators, many scholars also closely investigate the relationship between Treasury yield and stock return. The Fed model is widely used in literature papers to model the stock return by using Treasury yield. The Federal Reserve firstly adopted it in its Humphrey-Hawkins Report (1997), introduced a graph of the correlation between the S&P 500 earning-price ratio and Ten-year Treasury note yield. Antti Ilmanen (2003) examined the correlation between stock return and government bond return (treasury yield) in different time periods. He found that the correlation between the return of these two asset classes is positive through most of the 20th century. However, there are certain periods of time, like early 1930s and late 1950s, when these two variables are negatively correlated. He finally concluded that "growth and volatility shock" will enable stock return and treasury yield to move in the same direction, while "inflation shocks” tend to push these two to move in the opposite direction. This literature shows the general relationship between these two asset classes in a bigger picture.

Maikiel (2004) exploited the relationship between the stock return and the market price earning (P/E) multiples by using Fed model and Campbell-Shiller model. Although, he was convinced that stock valuation indicators such as dividend yields and market P/E multiples have the predictive power in the short-term and also have “considerable evidence of longer run negative serial correlation”, he claimed that there are no exploitable arbitrage opportunities for investors to take advantage on and gain excess return in the market.

Later, Maio (2013) examined the predictability of stock-bond yield, the difference between stock dividend yield and treasury bond yield. He found that the predictive power of the stock-bond yield gap will be strengthened with the absence of usual macro indicators, like dividend growth, etc. He created a dynamic accounting decomposition for the yield gap proxy, as a function of various other financial indicators, such as future short-term interest rate and future earnings growth, suggesting that the yield gap has a greater predictive power compared to other usual models in both short-term and long-term horizon.

Other asset classes, like commodity price, may also contain vital information about the future stock return. Many previous literatures studied the relationship between equity return and commodity price. Gorton and Rouwenhorst (2004) created an equally-weighted commodity future index and gathered data ranging from July of 1959 to March of 2004. They compared the return on commodity future and stock return as well as bond return. They finally concluded that commodity return is negatively correlated to the stock return and bond return due to the nature of business cycle. However, they found commodity is positively related to the inflation as well as the change in the expected future inflation.

Different from what discovered by Gorton, other scholars claimed a general positive correlation between stock return and commodity price. Black et al. (2014) researched the relationship between stock return and commodity price along with other indicators from 1973 to 2012. They picked SP500 index and Goldman Sachs Commodity Index (GSCI) index to represent stock performance and commodity price. Also, they used a regression model to formulate the connection and concluded that there is a long-run positive correlation between these two assets though in certain periods, the relationship between commodity price and stock return may vary.

More recently, Iyke and Ho (2021) further investigated the relationship between stock return and commodity price. Instead of using a commodity price index, they broke down that to each commodity individually with a larger sample space in more countries, Netherlands, the UK, and the US respectively. They used the autoregressive conditional heteroskedastic model (ARCH) to model the uncertainty with GLS estimation to formulate the relationship between stock return and commodity price. Their research showed commodity prices contained strong predictive power and "about 64% and 56% of the commodity returns can predict stock returns in-sample and out-of-sample, respectively".

What's more, not only the commodity price has been investigated with stock return, but also the volatility and its spillover effect as well as volatility transmission among these two asset classes has been studied by scholars. Mensi et al. (2013) researched "the return links and volatility transmission" between stock market and commodity market from 2000 to 2011 by using VAR-GARCH model. The model strongly indicated that there is volatility transmission among SP500 index and commodity market, especially gold and oil market. Also, the SP 500 index and gold index as well as the SP 500 index and WTI oil index have the greatest conditional correlation.

In contrast to the studies stated above, Huang et al. (1996) used vector autoregressive approach to investigate the relationship between oil future and the US stock return during 1980s. According to their study, “oil futures returns are not correlated with stock market returns, even contemporaneously, except in the case of oil company returns”. Moreover, some studies in other countries other than the United States also shows a similar conclusion regarding the relationship between stock return and commodity price. Nordin et al. (2020) examines the effect of commodity price, exchange rate and interest rate on Malaysian stock market performance. They use palm oil price, gold price and oil price to represent the commodity price. They employed the bound test approach and the result shows that “no significant influence was observed for both the oil price and gold price”. The relationship between stock return and commodity price in Malaysia may be insignificant.

As discussed above, treasury yield and commodity price can explain and forecast stock return in the future to a certain extent. This study will contribute to previous literature in the following ways. First, including treasury yield and commodity price into one model will better explain the variation in stock return and contain stronger predictive power. Second, the sample space will focus on the data of stock return and other variables in the recent ten years. After the quantitative easing and covid pandemic strike, the relationship between stock return and treasury yield along with commodity price may change compared to the conclusion in previous literature.

Methodology

In this paper, the relationship between stock return and treasury yields as well as commodity price will be determined. The model connecting stock return to treasury yield and commodity return should be:

$$SR_t = \alpha + \beta_1 Expansion_{t-1} TR_{t-1} + \beta_2 Expansion_{t-1} CR_{t-1} + \beta_3 Expansion_{t-1} IO_{t-1} + \beta_4 Expansion_{t-1} CB_{t-1} + \beta_5 Contraction_{t-1} TR_{t-1} + \beta_6 Contraction_{t-1} CR_{t-1} + \beta_7 Contraction_{t-1} IO_{t-1} + \beta_8 Contraction_{t-1} CB_{t-1} + \varepsilon_t$$

Equation 1

where SR, TR, CR, IO and CB stand for stock return, 10Y treasury bond return, commodity price index return, overnight interbank offered rate, and 10Y corporate bond return. t is the time subscript. ε_t is the error term associated with the model at time t. In order to deal with the problem of endogeneity within the model, according to Westerlund and Narayan (2012), the lag of variable should be included in the model. Expansion and Recession are dummy variables (Jacobsen et al., 2019). Purchasing Managers' Index (PMI) is a measure to indicate the trend of economic performance and it will be adopted to indicate whether the economy is expanding or not. When the PMI is greater than 50, the Expansion = 1 and Contraction = 0 and vice versa. Generalized least square estimation will be adopted to estimate the parameter of independent variables in stock return model and autoregressive conditional heteroskedastic (ARCH) model will be employed to model ε_t later, where:

$$\varepsilon_t = \sigma_{\varepsilon_t} z_t \qquad \text{Equation 2}$$

$$\sigma_{\varepsilon_t}^2 = \varphi_0 + \sum_{j=1}^{q} \varphi_j \varepsilon_{t-j}^2 \qquad \text{Equation 3}$$

where z_t stands for white noise existed in the market at time t; φ_s are constant and parameters before each lag of error term; q represents the optimal lag adopted in the model.

Partial autocorrelation function will be applied to determine the optimal lag number q in the ARCH model.

Data

All the time series data was extracted from the Wind Database. The closing price or yield for treasury bond, interbank offer rate and commodity index are highly explosive. Therefore, the returns of those variables are calculated to minimize the effect of self-correlation, but the Augmented Dickey Fuller test is still strongly against the alternative hypothesis that those time series data are stationary. The test results are shown in the following table:

Table 1:

Augmented Dickey Fuller Test results for time series data used

Category	Variable	H_0	H_A	P-value
US Model	Stock Return (SR)***	Explosive	Stationary	< 0.01
	10Y T-Bond Return (TR)***	Explosive	Stationary	< 0.01
	Commodity Return (CR)***	Explosive	Stationary	< 0.01
	Interbank Overnight Rate (IO)***	Explosive	Stationary	< 0.01
	10Y Corporate Bond Return (CB)***	Explosive	Stationary	< 0.01
	Expansion***	Explosive	Stationary	< 0.01
	Recession***	Explosive	Stationary	< 0.01
CN Model	Stock Return (SR)***	Explosive	Stationary	< 0.01
	10Y T-Bond Return (TR)***	Explosive	Stationary	< 0.01
	Commodity Return (CR)***	Explosive	Stationary	< 0.01
	Interbank Overnight Rate (IO)***	Explosive	Stationary	< 0.01
	10Y Corporate Bond Return (CB)***	Explosive	Stationary	< 0.01
	Expansion***	Explosive	Stationary	< 0.01
	Recession***	Explosive	Stationary	< 0.01

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively. Data is adapted from Wind Database.

The stock market return in the US and China will be represented by the return of S&P 500 index and CSI 300 index respectively. S&P 500 index is tracking the performance of the top 500 largest companies public listed in the United States; CSI 300 index is made up of top 300 stocks traded in Shanghai Stock Exchange and Shenzhen Stock Exchange. These two indexes can thoroughly illustrate the performance of the overall stock market in the two countries. Both countries' stock return is calculated as $SR_t = \ln(SP_t/SP_{t-1})$, where SP_t and SP_{t-1} stand for the closing price of the stock indexes in day t and day t-1. The return of the most active traded 10Y treasury bond in respective countries will be collected and they will be calculated as $TR_t = \ln(TY_t/TY_{t-1})$, where TP_t and TP_{t-1} are the closing yield of 10Y treasury bonds at day t and day t-1 respectively. Goldman Sachs Commodity Index (GSCI) and Wind Commodity Price Index will be used in the model to represent the performance of commodity in the US and China and both commodity return will be calculated as $CR_t = \ln(CP_t/CP_{t-1})$, where CP_t and CP_{t-1} are closing price of commodity price indexes at day t and day t-1. Liquidity indicator will also be adopted to reflect the effect of short-term liquidity on the overall stock market performance. USD London Interbank Offer Rate (LIBOR) overnight will demonstrate the overnight liquidity level in the US economy and Shanghai Interbank Offer Rate (SHIBOR) overnight will show that in the Chinese economy. This parameter in respective models will be both calculated as $IO_t = \ln(IOON_t/IOON_{t-1})$, where $IOON_t$ and $IOON_{t-1}$ represent the interbank offer rate overnight in two countries at day t and t-1. Finally, corporate bond return shows the performance of credit market which is high connected to the real economy. The rate of return in both economies will be calculated as $CB_t = ln(CBY_t/CBY_{t-1})$, where CBY_t and CBY_{t-1} stands for the closing corporate bond yield rate in day t and t-1 respectively. All of the data will be collected from January of 2012 to December of 2021.

The following table shows the summary of variables used to model the return of the Chinese stock market and the US stock market:

Table 2:

Summary of variables (Modeling US stock return)

Variables	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Stock Return (SR)	-0.127652	-0.003211	0.000690	0.000524	0.005186	0.089683
10Y T-Bond Return (TR)	-0.315080	-0.013990	0.000000	-0.000020	0.013890	0.341750
Commodity Return (CR)	-0.125222	-0.006289	0.000588	-0.000240	0.006541	0.076167
Interbank Overnight Rate (IO)	0.050750	0.092500	0.153100	0.636360	1.178330	2.402750
10Y Corporate Bond Return (CB)	-0.100083	-0.008535	0.000000	-0.000045	0.007782	0.242072
Expansion	0.000000	1.000000	1.000000	0.884300	1.000000	1.000000
Recession	0.000000	0.000000	0.000000	0.115700	0.000000	1.000000

Source: Wind Database

Table 3:

Summary of variables (Modeling Chinese stock return)

Variables	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Stock Return (SR)	-0.091542	-0.005987	0.000398	0.000315	0.007130	0.064989
10Y T-Bond Return (TR)	-0.058508	-0.003892	0.000000	-0.000088	0.003822	0.041367
Commodity Return (CR)	-0.052607	-0.005353	0.000137	0.000013	0.005608	0.053081
Interbank Overnight Rate (IO)	0.602000	1.968000	2.315000	2.379000	2.693000	13.444000
10Y Corporate Bond Return (CB)	-0.046617	-0.002283	0.000000	-0.000138	0.001958	0.034227
Expansion	0.000000	1.000000	1.000000	0.803800	1.000000	1.000000
Recession	0.000000	0.000000	0.000000	0.196200	0.000000	1.000000

Source: Wind Database

Discussion

By using the GLS regression model, the parameters of variables are estimated and models for the US and Chinese stock return are shown in the following table:

Table 4:

US Stock Return Model

Parameter	**Value**	**Standard Error**	**P-value**
Intercept***	0.00059691	0.00017852	8.3966×10^{-4}
$Exp \times TR$***	0.17652613	0.01035780	1.5355×10^{-61}
$Exp \times CB$***	-0.15792371	0.01720806	9.3186×10^{-20}
$Exp \times CR$***	0.23438317	0.01648334	4.2305×10^{-44}
$Exp \times IO$***	0.01346172	0.00436290	2.0555×10^{-3}
$Con \times TR$***	0.19822633	0.01904420	7.6606×10^{-25}
$Con \times CB$***	-0.20603080	0.04130240	6.5298×10^{-7}
$Con \times CR$***	0.17396403	0.02998729	7.4575×10^{-9}
$Con \times IO$	-0.01112881	0.00812884	1.7111×10^{-1}

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively

Table 5:

Chinese Stock Return Model

Parameter	Value	Standard Error	P-value
Intercept	0.0003245	0.00028074	2.4781×10^{-1}
$Exp \times TR$***	0.1763696	0.04534133	1.0302×10^{-4}
$Exp \times CB$	-0.0888676	0.07018358	2.0556×10^{-1}
$Exp \times CR$***	0.3357788	0.03114717	1.6850×10^{-26}
$Exp \times IO$	-0.0034842	0.00309115	2.5979×10^{-1}
$Con \times TR$***	0.2650029	0.08476867	1.7919×10^{-3}
$Con \times CB$	-0.1930682	0.15554245	2.1463×10^{-1}
$Con \times CR$***	0.4039038	0.06124509	5.2080×10^{-11}
$Con \times IO$	-0.0016762	0.00684993	8.0670×10^{-1}

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively

In the US stock return model, all the parameters have a p-value that is lower than 0.01 which indicates the significance of those estimated parameters and variables in the model, except variable Interbank Overnight Rate during contractionary period. Therefore, the variable IO during the contractionary period is dropped from the preliminary GLS US stock return model.

The estimated values of parameters are also very reasonable. In both expansionary and contractionary period, the daily return of 10Y US Treasury Bond yield and GSCI contribute positively to the return of SP500 index. On the other hand, the daily return of 10Y US corporate bond yield seems to contribute negatively to the return of SP500 index. These relationships are highly in line with economic theories and facts that:

1. Stocks and Treasury Bonds are two substitutable asset classes and the increase in demand of one asset will inevitably decrease the demand of the other one. Therefore,

stock return and yield of Treasury Bond should move in the same direction.

2. The increasing demand of commodity, leading to increasing price of commodity, will indicate the expansion of the economy and therefore positively impact the stock return.
3. The drop in the yield of corporate bond reflects the rising demand of corporate bond. This phenomenon is usually due to the improving fundamental of US corporations and thus result in the rise of stock index.

Moreover, the Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) of the US stock return model are -15787.06 and -15729.32, respectively. The small value of these two criteria indicates the sophistication of the model. Thus, by substituting parameters into Equation 1, the final GLS model for the SP 500 index return is:

$$SR_t = 0.0005969 + 0.1765 Expansion_{t-1} TR_{t-1} + 0.2344 Expansion_{t-1} CR_{t-1} + 0.01346 Expansion_{t-1} IO_{t-1} - 0.1579 Expansion_{t-1} CB_{t-1} + 0.1982 Contraction_{t-1} TR_{t-1} + 0.1740 Contraction_{t-1} CR_{t-1} - 0.2060 Contraction_{t-1} CB_{t-1} + \varepsilon_t \quad \text{Equation 4}$$

In contrary to the US stock return model, some variables seem to underperform in the Chinese capital market. In Table 4, only the return of 10Y China Treasury Bond yield and commodity index in all time can reasonably explain the variation of the market with p-values lower than 0.05. The return of 10Y corporate bond yield and interbank overnight offer rate are less significant in modeling the return of CSI 300 index in both expansionary and contractionary period. Therefore, these two variables may be dropped from the preliminary GLS model for CSI 300 index.

The return of 10Y China Treasury Bond yield and commodity index are positively correlated to the return of CSI 300 Index. These relationships are the same as those in the US stock return model, indicating the universality of these two variables used to modeling the stock market return. What's more, The AIC and BIC of the model are -13846.7 and -13788.78, which are higher than those of the US stock return model, indicating less variation of the stock return interpretated by the same variables in Chinese capital market. Therefore, according to the regression result and variable significance check, the GLS model of Chinese stock return is:

$$SR_t = 0.0003245 + 0.1764 Expansion_{t-1} TR_{t-1} + 0.3358 Expansion_{t-1} CR_{t-1} + 0.2650 Contraction_{t-1} TR_{t-1} + 0.4039 Contraction_{t-1} CR_{t-1} + \varepsilon_t \quad \text{Equation 5}$$

By plotting the residuals from both models, these two time series residuals suggest that more variations are unexplained in Chinese stock market than those in the US stock market:

Figure 1:

US stock return model residuals

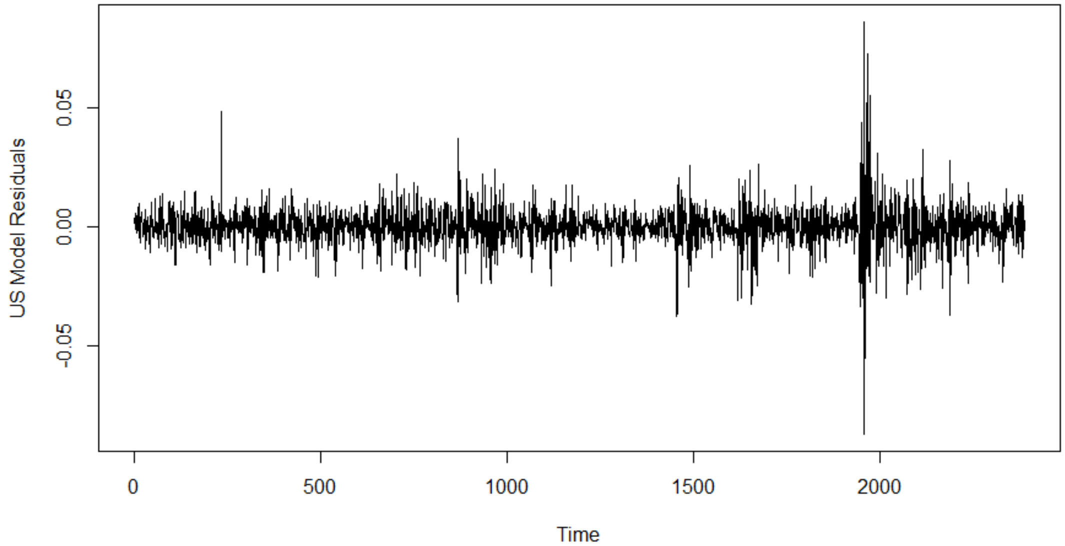
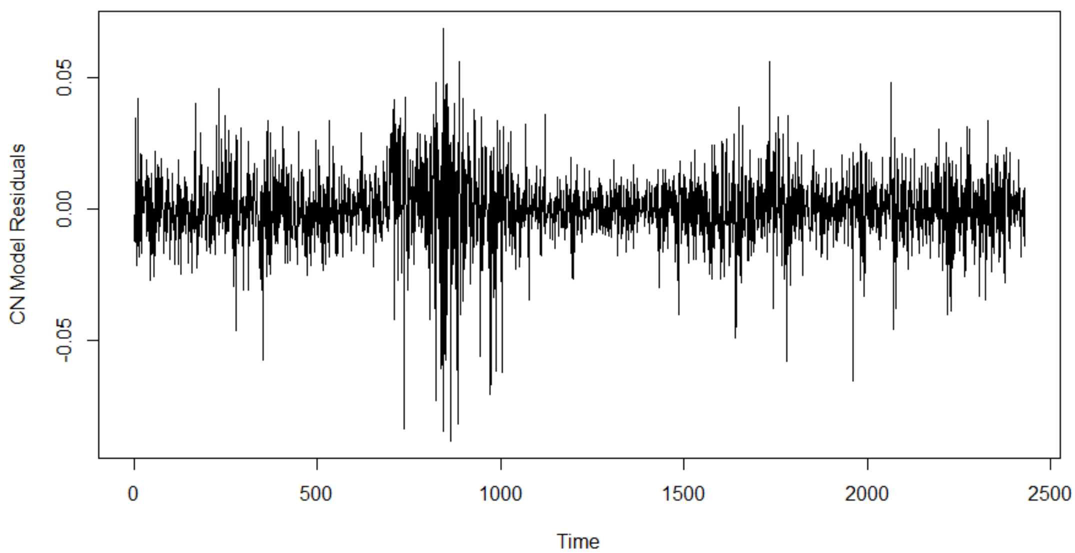


Figure 2:

China stock return model residuals



In both figures, residuals are centered in 0 with variation up and down. By comparing two time series residuals generated from the models, the Chinese stock return model residuals are more volatile in most of the time. Also, there is a common characteristic embedded in the figures: high residuals clustered in some certain period of time and low residuals dominate the other periods. Therefore, autoregressive conditional heteroskedastic (ARCH) model can be employed to model ε_t in Chinese stock return and US stock return. The Lagrange Multiplier (LM) test for ARCH is employed and the results are:

Table 6:

LM ARCH Test Result for residuals in US and CN model

Residuals	$\boldsymbol{H_0}$	$\boldsymbol{H_A}$	**P-value**
Residuals in US model***	No ARCH effect	ARCH effect exists	$< 2.2 \times 10^{-16}$
Residuals in CN model***	No ARCH effect	ARCH effect exists	$< 2.2 \times 10^{-16}$

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively

Both tests for ARCH effect have p-value under 0.01 and the null hypothesis is rejected. Therefore, both residuals from the two GLS models can be modeled by ARCH method based on Equation 2 and Equation 3.

Figure 3:

PACF of ε^2 in US stock return model

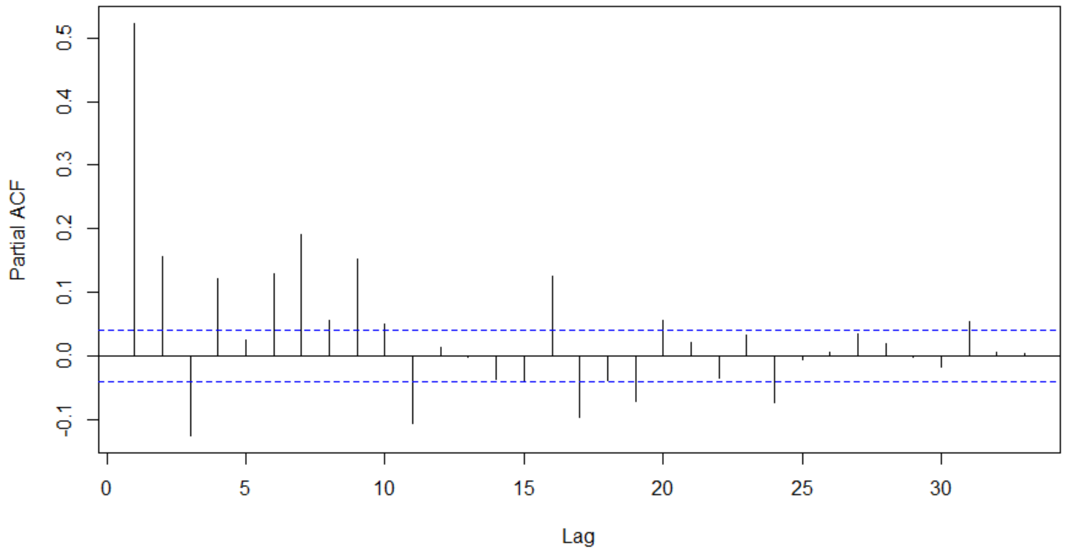
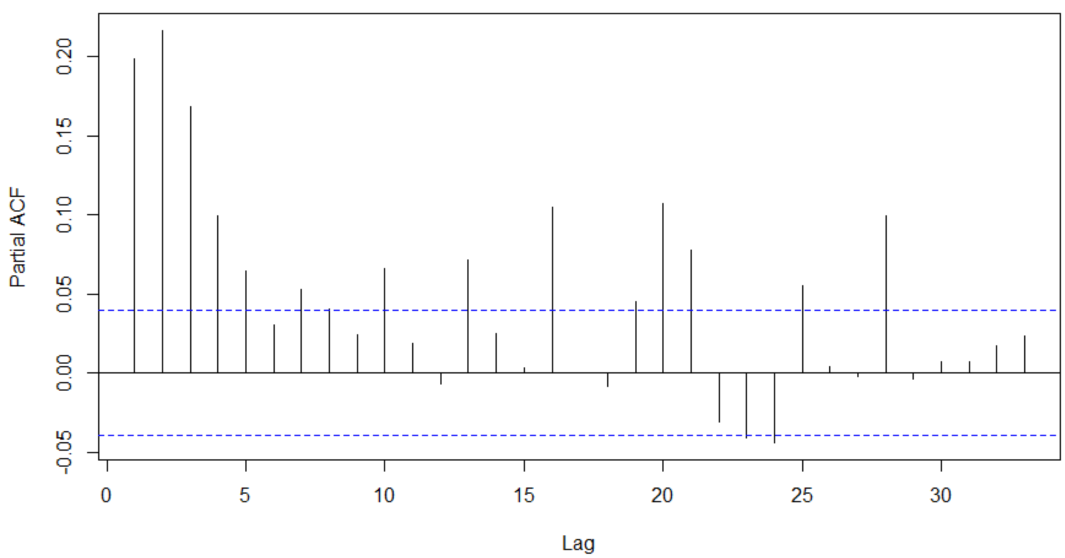


Figure 4:

PACF of ε^2 in Chinese stock return model



From the two figures about the lag of ε^2 in respective models, the first 11 lags of ε^2 is out of the confidence interval in the US stock return model and the first 7 lags of ε^2 is out of bound in the Chinese stock return model. Therefore, initially q_{US} (optimal lags in the US stock return residual ARCH model) is set to 11 and q_{CN} (optimal lags in the Chinese stock return residual ARCH model) is set to 7. The ARCH model results in:

Table 7:

ARCH result for ε^2 in US stock return model (q = 11)

Lags	**Parameter value**	**P-value**
Intercept***	2.001×10^{-5}	$< 2.00 \times 10^{-16}$
Lag 1***	1.232×10^{-1}	4.00×10^{-15}
Lag 2***	9.333×10^{-2}	4.38×10^{-7}
Lag 3***	8.967×10^{-2}	1.18×10^{-7}
Lag 4***	7.805×10^{-2}	9.21×10^{-5}
Lag 5***	5.051×10^{-2}	0.00209
Lag 6***	5.918×10^{-2}	0.00178
Lag 7***	4.264×10^{-2}	0.00243
Lag 8	1.695×10^{-14}	1.00000
Lag 9	2.112×10^{-2}	0.15535
Lag 10***	5.330×10^{-2}	0.00218
Lag 11	1.872×10^{-2}	0.21660

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively

Table 8:

ARCH result for ε^2 in Chinese stock return model (q = 7)

Lags	Parameter value	P-value
Intercept***	5.866×10^{-5}	$< 2.00 \times 10^{-16}$
Lag 1***	7.327×10^{-2}	1.90×10^{-7}
Lag 2***	9.802×10^{-2}	8.25×10^{-9}
Lag 3***	1.067×10^{-1}	2.85×10^{-7}
Lag 4***	9.912×10^{-2}	7.55×10^{-15}
Lag 5***	8.356×10^{-2}	1.88×10^{-7}
Lag 6***	1.131×10^{-1}	2.22×10^{-16}
Lag 7***	1.504×10^{-1}	$< 2.00 \times 10^{-16}$

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively

The ARCH result of 7 lags for ε^2 in Chinese stock return model is promising since all the p-values of lags parameter are below 0.01. However, the result for US stock return model is not very convincing. Before lag 8, the p-values of parameters are all below 0.01, but starting from lag 8, the significance of parameters may be diminishing. Therefore, the optimal lag of ε^2 in US stock return model will also be 7 and the updated ARCH result is:

Table 9:

Updated ARCH result for ε^2 in US stock return model (q = 7)

Lags	**Parameter value**	**P-value**
Intercept***	1.860×10^{-5}	$< 2.00 \times 10^{-16}$
Lag 1***	2.100×10^{-1}	$< 2.00 \times 10^{-16}$
Lag 2***	1.291×10^{-1}	2.66×10^{-8}
Lag 3***	1.383×10^{-1}	2.40×10^{-10}
Lag 4***	1.006×10^{-1}	5.76×10^{-6}
Lag 5***	4.218×10^{-2}	0.007636
Lag 6***	7.134×10^{-2}	5.13×10^{-5}
Lag 7***	5.670×10^{-2}	0.000389

Note: ***, **, * represents significant level of 1%, 5% and 10% respectively

The variance of error term of the US and Chinese stock return model can be plotted based on the result of their ARCH results:

Figure 5:

Variance of error term in the US stock return model based on ARCH (7) result

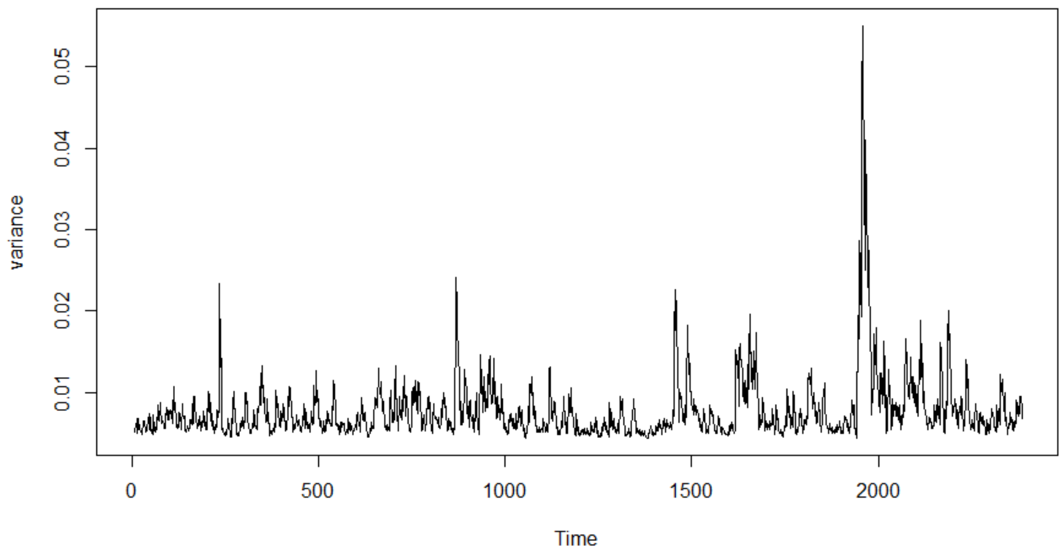
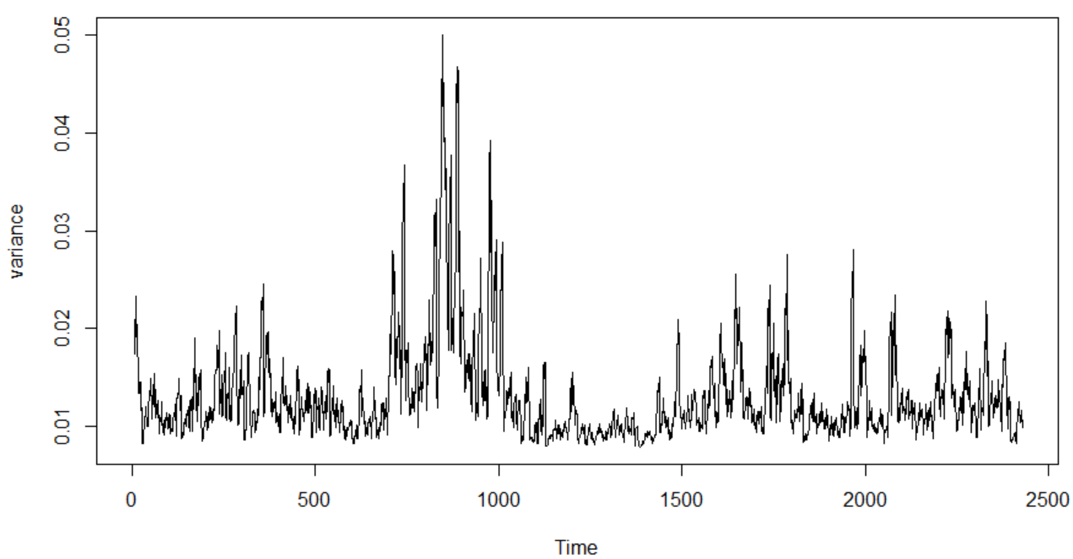


Figure 6:

Variance of error term n the Chinese stock return model based on ARCH (7) result



The two figures above show the variance of error terms, or the unexplainable part of the market under the GLS model derived earlier, embedded in the respective models. This ARCH (7) model generates the predicted variance that is close to real market. In the graph about residuals from US stock return model, the unusual peak occurs in 2020 represents the market turmoil happened in 2020 March period when coronavirus became prevalent in the US. In the other graph, the greatest cluster of high variances occurred in 2014 to 2015 and at that time due to the prevalence of high leverage used by individual investors, the CSI 300 index increased more than 150% from mid of 2014 to mid of 2015 and then decreased about 40% due to the burst of the leverage bubble.

Generally, by comparing two graphs, it also reveals that the variance of error terms in the US stock return model is lower than that in Chinese stock return model. In another word, the overall market risk in China is higher than that in the US and the return of the index is more volatile in China. This finding is aligned with the fact that Chinese capital market is more immature than the US capital market in terms of the imperfection of fundamental system design and higher number of individual investors, which drastically amplifies the randomness and volatility in the stock market.

The modeling of error term can not only help to improve the preciseness of the GLS model but also help investors to foresee the upcoming high volatility period of the stock market. As a result, it may help investors to lower their leverage or close their position to be ready for the market turmoil.

The error term in two models can be formulated as below based on ARCH (7) results based in Equation 2:

$$\varepsilon_{US\_t} = \sqrt{0.0000186 + 0.210\varepsilon_{t-1}^2 + 0.129\varepsilon_{t-2}^2 + 0.138\varepsilon_{t-3}^2 + 0.101\varepsilon_{t-4}^2 + 0.0422\varepsilon_{t-5}^2 + 0.0713\varepsilon_{t-6}^2 + 0.0567\varepsilon_{t-7}^2}\, z_t$$

Equation 6

$$\varepsilon_{CN\_t} = \sqrt{0.0000587 + 0.0733\varepsilon_{t-1}^2 + 0.0980\varepsilon_{t-2}^2 + 0.107\varepsilon_{t-3}^2 + 0.0991\varepsilon_{t-4}^2 + 0.0836\varepsilon_{t-5}^2 + 0.113\varepsilon_{t-6}^2 + 0.150\varepsilon_{t-7}^2}\, z_t$$

Equation 7

where z_t is the white noise process that is unable to be modeled.

Finally, the GLS model can be improved based on further modeling the error terms. By substituting Equation 6 and 7 into Equation 4 and 5 respectively, the finalized GLS models for US stock return and Chinese stock return are:

$$SR_{US\_t} = 0.0005969 + 0.1765Expansion_{t-1}TR_{t-1} + 0.2344Expansion_{t-1}CR_{t-1} + 0.01346Expansion_{t-1}IO_{t-1} - 0.1579Expansion_{t-1}CB_{t-1} + 0.1982Contraction_{t-1}TR_{t-1} + 0.1740Contraction_{t-1}CR_{t-1} - 0.2060Contraction_{t-1}CB_{t-1} + \sqrt{0.0000186 + 0.210\varepsilon_{t-1}^2 + 0.129\varepsilon_{t-2}^2 + 0.138\varepsilon_{t-3}^2 + 0.101\varepsilon_{t-4}^2 + 0.0422\varepsilon_{t-5}^2 + 0.0713\varepsilon_{t-6}^2 + 0.0567\varepsilon_{t-7}^2}\, z_t$$

Equation 8

$$SR_{CN\_t} = 0.0003245 + 0.1764Expansion_{t-1}TR_{t-1} + 0.3358Expansion_{t-1}CR_{t-1} + 0.2650Contraction_{t-1}TR_{t-1} + 0.4039Contraction_{t-1}CR_{t-1} + \sqrt{0.0000587 + 0.0733\varepsilon_{t-1}^2 + 0.0980\varepsilon_{t-2}^2 + 0.107\varepsilon_{t-3}^2 + 0.0991\varepsilon_{t-4}^2 + 0.0836\varepsilon_{t-5}^2 + 0.113\varepsilon_{t-6}^2 + 0.150\varepsilon_{t-7}^2}\, z_t$$

Equation 9

Conclusion

This paper focused on modeling the stock return in China and US respectively from the beginning of 2012 to the end of 2021. Based on previous literature, 10Y Treasury bond yield, 10Y corporate bond yield, commodity price index and overnight interbank offer rate are utilized to model the return of stock index, CSI 300 index in China and SP 500 index in the US. Also, dummy variables Expansion and Contraction are adopted to better model the stock return in different phase of economic cycle. Generalized Least Square model is used to generate the preliminary model for the stock return in two countries. In the US stock market, the overnight interbank offer rate, USD LIBOR ON, is less significant during the contractionary period comparing to other variables. In the Chinese stock market, 10Y corporate bond yield and overnight interbank offer rate, SHIBOR ON, are not significant in both expansionary and contractionary economic period. The US stock return model also has a higher overall ability to explain the variation of the stock return given its lower AIC and BIC value compared with those of the Chinese stock return model. Moreover, the error term associated with two models is considered to have autoregressive conditional heteroskedastic (ARCH) effect, backed by the Lagrange Multiplier (LM) test for ARCH effect. The optimal lag is determined by the Partial Autocorrelation Function and the significance of parameter estimated. The result shows that the residuals from US stock return model and Chinese stock return model can be both formulated by ARCH (7) model. The ARCH model successfully formulates the relationship among lagged error terms from the GLS model and anticipate the volatile market period of the US stock market in 2020 due to the hit of coronavirus and the Chinese stock market in 2014 and 2015 due to the leverage bubble. Besides, the modeling of error terms and the stock return also reveals the immaturity of the Chinese stock market compared to the US stock market. Finally, the successful model of the stock return and the variance of the unexplained part of the market variation suggested in this paper will largely help investor to manage their stock position and the level of leverage by carefully watch the level of variance of the error term, the unexplained market risk, associated with the GLS models in both countries.

Reference

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, *20*(3), 651–707. https://doi.org/10.3386/w8207

Black, A. J., Klinkowska, O., McMillan, D. G., & McMillan, F. J. (2014). Forecasting stock returns: Do commodity prices help? *Journal of Forecasting*, *33*(8), 627–639. https://doi.org/10.1002/for.2314

Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, *21*(4), 1509–1531. https://doi.org/10.1093/rfs/hhm055

The Federal Reserve Board. (1997). (rep.). *Humphrey-Hawkins Report*. Retrieved November 17, 2021, from https://www.federalreserve.gov/boarddocs/hh/1997/july/ReportSection2.htm.

Gorton, G., & Rouwenhorst, K. G. (2004). (working paper). *Facts and Fantasies about Commodity Futures*. Retrieved November 17, 2021, from https://www.nber.org/papers/w10595.

Huang, R. D., Masulis, R. W., & Stoll, H. R. (1996). Energy shocks and Financial Markets. *Journal of Futures Markets*, *16*(1), 1–27. https://doi.org/10.1002/(sici)1096-9934(199602)16:1<1::aid-fut1>3.0.co;2-q

Ilmanen, A. (2003). Stock-bond correlations. *The Journal of Fixed Income*, *13*(2), 55–66. https://doi.org/10.3905/jfi.2003.319353

Iyke, B. N., & Ho, S.-Y. (2021). Stock return predictability over four centuries: The role of Commodity Returns. *Finance Research Letters*, *40*. https://doi.org/10.1016/j.frl.2020.101711

Jacobsen, B., Marshall, B. R., & Visaltanachoti, N. (2019). Stock market predictability and industrial metal returns. *Management Science*, *65*(7), 3026–3042. https://doi.org/10.1287/mnsc.2017.2933

Kolchin, K., Podziemska, J., & Mostafa, A. (2021). (rep.). *Capital Markets Fact Book, 2021*. SIFMA. Retrieved November 17, 2021, from https://www.sifma.org/resources/research/fact-book/.

Maio, P. (2013). The “Fed model” and the predictability of stock returns*. *Review of Finance*, *17*(4), 1489–1533. https://doi.org/10.1093/rof/rfs025

Malkiel, B. G. (2004). Models of stock market predictability. *Journal of Financial Research*, *27*(4), 449–459. https://doi.org/10.1111/j.1475-6803.2004.00102.x

Mensi, W., Beljid, M., Boubaker, A., & Managi, S. (2013). Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. *Economic Modelling*, *32*, 15–22. https://doi.org/10.1016/j.econmod.2013.01.023

Nordin, N., Nordin, S., & Ismail, R. (2020). The impact of commodity prices, interest rate and exchange rate on stock market performance: An empirical analysis from Malaysia. *Malaysian Management Journal*. https://doi.org/10.32890/mmj.18.2014.9015

Westerlund, J., & Narayan, P. K. (2012). Does the choice of estimator matter when forecasting returns? *Journal of Banking & Finance*, *36*(9), 2632–2640. https://doi.org/10.1016/j.jbankfin.2012.06.005

World Bank. (2021). *Market capitalization of listed domestic companies (% of GDP)*. Data. Retrieved November 17, 2021, from https://data.worldbank.org/indicator/CM.MKT.LCAP.GD.ZS?end=2020&most_recent_value_desc=true&start=1975.

World Bank. (2021). *Market capitalization of listed domestic companies (current US$)*. Data. Retrieved November 17, 2021, from https://data.worldbank.org/indicator/CM.MKT.LCAP.CD?end=2020&most_recent_value_desc=true&start=1975.