|  |  |  |
| --- | --- | --- |
| **Problem Chosen** C | **2022 MCM/ICM Summary Sheet** | **Team Control Number** 2202151 |

title

Abstract

**Key Words**:

Content

# 1. Introduction

1.1 Background

When comes to gold and bitcoin, traders buy and sell them frequently under market regime due to gaining maximum return. In addition, traders are supposed to take commissions for purchase and sale into consideration.

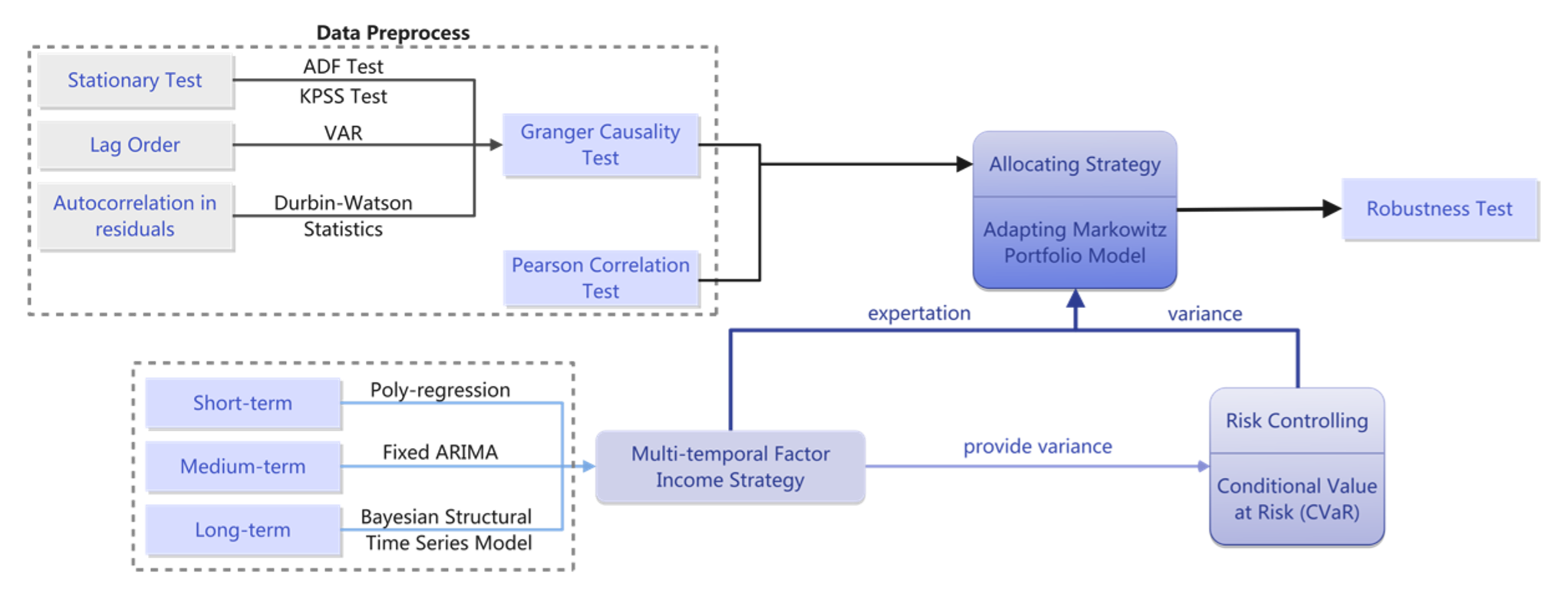
1.2 Our work

For the sake of locating the best portfolio in five-year trading period, we formulate the specific trading schedule of gold and bitcoin from 9/11/2016 to 2021/9/10 with the initial 1,000 dollars investment. Furthermore, the commission for each transaction (purchase or sale) costs α% of the amount traded and we assume that αgold = 1% and αbitcoin = 2%. It’s worth noticing that only can gold transaction take place on days the market is open, while bitcoin transaction can be traded every day.

We will proceed as follows for the sake of tackling these problems:

* Build a model to find out the optimal daily trading strategy on the basis of given data up to that day. We are also required to calculate the return on 9/10/2021 via our model and strategy. Subsequently, prove that the above approach can yield the maximum return.
* Determine the interaction between transaction costs and strategy and how do transaction costs affect results.
* Write a memorandum which involves our strategy, model and results.

The whole modeling process can be shown as follows:



**Fig.1** Technology route for the creation of our paper.

# 2. Assumptions and Justification

To simplify the given problems and modify it more appropriate for simulating real-life conditions, we make the following basic hypotheses, each of which is properly justified.

* We assume that the closing price is representative of the day and do not take into account price fluctuations during the day. This assumption is the premise for our in-depth study. For virtual currencies such as Bitcoin, there is no up and down constraint, no trading time limit, and great volatility can occur from time to time in a short period of time.
* We assume that all investors react to price changes in a timely and rational manner and do not cause large price fluctuations due to blind acquisitions or sell-offs due to emotions.

# 3. Notations

We list the symbols and notations used in this paper in Table 1.

**Table 1 Notations**

|  |  |
| --- | --- |
| Symbols | Definition |
|  | Economic recession index |
|  | Ecosystem sustainability index |
|  | Social habitability index |
|  | Fragility index based on the climate change |
|  | Climate change index |
|  | Pearson’s contingency coefficient |
|  | Total cost of human intervention |
| , | The Gaussian error |
| , | The structural parameters |
|  | The vector of expected returns of the investments |
|  | The covariance matrix of the investments |
|  | The allocation weight of the investments in the portfolio |

# 4. Data Preprocessing

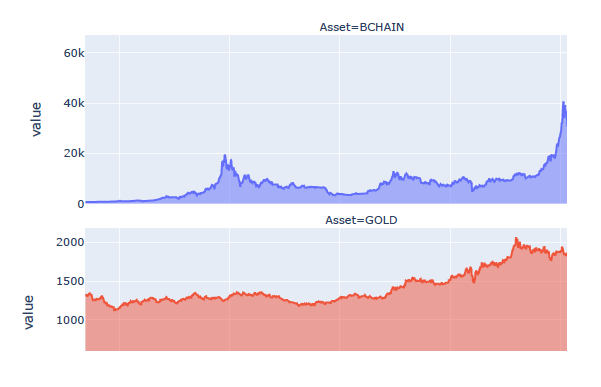
In this section, we preprocess the data set and analyze the trading relationship between bitcoin and gold.

**4.1 Data visualization**

Due to the arrangement of different trading schedule of gold and bitcoin, we delete partial dates and corresponding closing prices of bitcoin transaction on the basis of the rules of gold trading regime, i.e., cutting out the information of bitcoin on gold’s odd dates.

By observing the data set of gold daily prices, we realize that there exist two missing values on certain days. On account of abundant five-year data, we ignore the two days’ information of gold transaction, as well as bitcoin.

Subsequently, we visualize the modified data as follows:

****

**Fig 2 Bitcoin and Gold daily prices**

**4.2 Granger Causality Test**

In this subsection, we discuss the causal relationship between gold and bitcoin prices. Simultaneously, associating with actual investment, we are aware that only past events can affect the present and future events, while the present and future events won’t influence on past events. For example, if we are trying to explore whether the variable gold has a causal effect on the variable bitcoin, then we are only required to estimate if the lag of gold affects the present value of bitcoin. As a result, we analysis the cause and effect between the two elements via Granger Causality Test, which specifically used to test the reason why one set of time series X is the cause of another set of time series Y. So, if we would control for the past value of bitcoin, the past value of gold can still have significant explanatory power for the variable bitcoin and we record that gold has Granger-cause on bitcoin.

**4.2.1 ADF Test and KPSS Test for Stationary**

On the account of Granger Causality Test’s demand for stationarity, we perform stationarity analysis on the data set though augmented Dickey–Fuller test (ADF) tests and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test.

The ADF test can be used to help us understand whether the timeseries is stationary or not. The KPSS test figures out if a time series is stationary around a mean or linear trend, or is non-stationary due toa unit root. We utilize **Null hypothesis** and **Alternative hypothesis**, which represent the timeseries is not stationary and stationary in ADF test and stand for the opposite of ADF in KPSS, to describe the stationarity of the time series.

After cross-checking ADF test and KPSS test, we gain the p-values of gold and bitcoin and exhibit them in table 2.

**Table 2 取名**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| p-value  product | ADF | KPSS | ADF  (first order difference) | KPSS  (first order difference) |
| gold | 0.957633 | 0.01 | 0.000000 | 0.1 |
| bitcoin | 0.894157 | 0.01 | 0.000000 | 0.1 |

According to table 2, we find out that for ADF test, when the p-values are all well above the 0.05 alpha level, we cannot reject the null hypothesis. Hence the two timeseries are not stationary. For KPSS test, the p-value are all less than 0.05 alpha level, therefore, we can reject the null hypothesis and derive that the two timeseries are not stationary.

Subsequently, we transform the timeseries to be stationary by **first order difference** method and we present the results in table 2. Obviously, the outcomes satisfy Alternative hypothesis. Then we adopt the results after first order difference to conduct Granger Causality test.

**4.2.2 VAR Model**

Since the data set only includes trading dates and corresponding price, it’s difficult for us to calculate the causal correlation coefficient between gold and bitcoin. So, we try to build the model with several **lag period**. Here we apply Vector Autoregression (VAR), which is a statistical model used to capture the relationship between quantities as they change over time. Besides, for the reason that the VAR model assumes that the passed time series are stationary, we take advantage of data after first order difference due to ADF test and KPSS test.

Though VAR model, we generalize the single-variable autoregressive model by allowing for multivariate time series. The VAR model describes that n variables (endogenous variables) within the same sample period can be linear functions of their past values. The p-value reduced-form VAR formula is as follows:

where represent the predicted price of gold/bitcoin. The variables of the form indicate that variable's value time periods earlier and are called the "th lag" of . The variable is a -vector of constants serving as the intercept of the model. is a time-invariant (*k × k*)-matrix and et is a *k*-vector of error terms. The error terms must satisfy the following conditions: . i.e., every error term has a mean of zero.

In fact, there is no hard-and-fast-rule on the choice of lag order. In this paper, we use the AIC in selecting the lag order with the smallest value. When lag order is an integer between 1 and 15, the corresponding AIC is small, about 18.23. Therefore, we will select lag order =15, i.e., we apply the former 15 transactions from the current trading date to predict the amount. In the following statement, we fit the correlation coefficient via order 15 linear function. Meanwhile, we obtain the correlation of residuals between bitcoin and gold is 0.007169, which means there is little correlation.

**4.2.3 Durbin-Watson Statistics for Residual Autocorrelation Test**

After determining the small correlation of residuals between bitcoin and gold, we examine the autocorrelation of residuals by Durbin-Watson statistics. If is the residual given by , the Durbin-Watson statistic states that null hypothesis: , alternative hypothesis: , then the test statistic is

where is the number of observations.

Since is approximately equal to 2, where is the sample autocorrelation of the residuals, = 2 indicates no autocorrelation. The value of always lies between 0 and 4. If the Durbin–Watson statistic is substantially less than 2, there is evidence of positive serial correlation. If > 2, successive error terms are negatively correlated. In regressions, this can imply an underestimation of the level of statistical significance.

By computation, we attain the autocorrelation coefficients of gold and bitcoin equal to 2.0 and 1.99, respectively. As a result, there is no autocorrelation detected in the residuals.

**4.2.4 Granger Causality Test Results**

For illustration, consider a bivariate linear autoregressive model of two variables  and :

where is the maximum number of lagged observations included in the model, the matrix contains the coefficients of the model (i.e., the contributions of each lagged observation to the predicted values of and , and and are residuals for each time series.

In the end, under the circumstance that 15 lag order are selected and the known residual terms have no autocorrelation, we gain the Granger causation matrix in table 3.

**Table 3** Granger causation matrix of gold and bitcoin

|  |  |  |
| --- | --- | --- |
|  | Gold | Bitcoin |
| Gold | 1.0000 | 0.3720 |
| Bitcoin | 0.0255 | 1.0000 |

As we can see in Table 3, the results of Granger causality test are relatively small. Hence, we conclude that there is no obvious cause and effect between gold and bitcoin transaction.

**4.3 Pearson Correlation Test**

In addition to examining the casual relationship between gold and bitcoin, we also evaluate the correlation between them. We adopt Pearson correlation coefficient to determine whether gold and bitcoin are related. We denote Pearson correlation coefficient between X and Y by and its formula is as follows:

where represents covariance of X and B. andindicate the expectation of X and Y, respectively.

First of all, we discuss the relationship between gold and bitcoin over time. Using Pearson correlation coefficient formula, we realize there is no significant correlation between gold and bitcoin since the Pearson Correlation Coefficient is 0.6492929578934703 with a P-value of P = 5.528011709457134410-219.

Next, we take time ductility of events into consideration and further calculate that the Pearson correlation coefficient reached its maximum value at 229 days, which was 0.8808. Apparently, the time interval of interaction between gold and bitcoin can’t last such a long period of time.

# 5. Adapting Markowitz Pair-Portfolio

**5.1 Portfolio Strategy**

Markowitz's modern asset allocation theory, one frequently adopted Portfolio strategy, takes as input the expected return as exception and covariance matrix of these investments as risk. It also finds the optimal portfolio by optimizing the following objective function.

where denotes the vector of expected returns of the investments, denotes the covariance matrix of the investments, and is the allocation weight of the investments in the portfolio.

the optimal solution, that is, the best allocation of assets is：

This model is criticized by professional investment institutions in practice is mainly because its input is very strict: investors must provide the expected return and covariance of the investment products to be allocated. Once the predicted value is very outrageous, then the maximization of asset allocation utility becomes the maximization of error. Chopra and Ziemba (1993) Chopra, V. K. and W. T. Ziemba (1993). The effort of errors in means, variances, and covariances on optimal [portfolio choice](https://www.zhihu.com/search?q=portfolio+choice&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22article%22%2C%22sourceId%22%3A%2238282835%22%7D) Journal of Portfolio Management, Vol. 19(2), 6 – 11. pointed out that the error of return expectation has an order of magnitude higher impact on asset allocation than the impact of covariance.

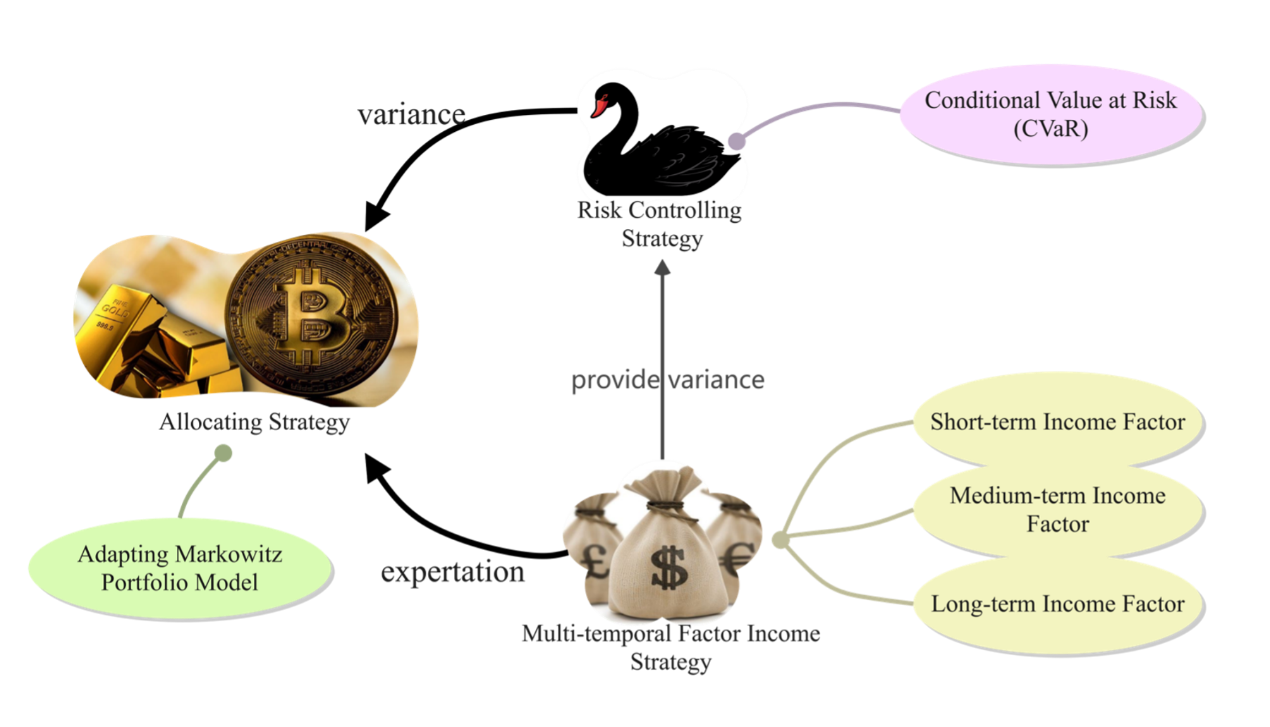
Later, many professional quantitative investment researchers made improvements based on the Markowitz model, such as Black-Litterman's correction of returns under the Bayesian framework. Inspired by the previous work and considering the limited data available, we redefine the Markowitz model based on the forecasting from incoming model and risk calculated by Conditional Value at Risk model:

Support the exception and Variance of gold short, medium and long-term factors respectively are：

and we have the following formulas:

so, we get and .

then we will introduce the forecasting by Multi-temporal Factor Incoming Strategy as exception and the risk by Risk Controlling through Conditional Value at Risk (*CVaR*) as variance.



**Fig 3 Adapting Markowitz Portfolio Model**

**5.2 Multi-temporal Factor Income Strategy**

In most quant investment strategies, income model includes several factors to measure the product’s the ability of earnings like Fama-French Model containing wave, trend and risk rate. But the data available, history price merely, is limited compared to that required in Fama-French Model. Correspondingly, we estimate three factors based on time series of different lengths including short, medium, and long term income factors. Their different characteristics are shown in Table 4:

**Table 4** Granger causation matrix of gold and bitcoin

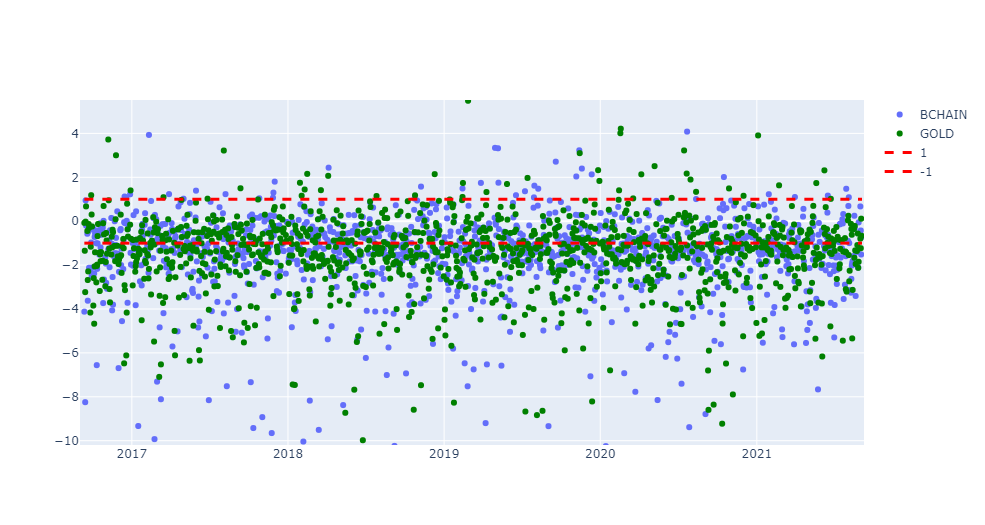
|  |  |  |  |
| --- | --- | --- | --- |
| **Factor Name** | **Model** | **Characteristic** |  |
| Short Term | Poly-Regression | Fluctuation | 5 |
| Medium Term | Fixed ARIMA | Trend | 20 |
| Long Term | Bayesian Structural Time Series | Pattern Analysis | 200 |

where indicates the sliding window length in model forecasting.

Then, we state the details of different term income factor.

**5.2.1 Short term Income Factor based on Poly-regression**

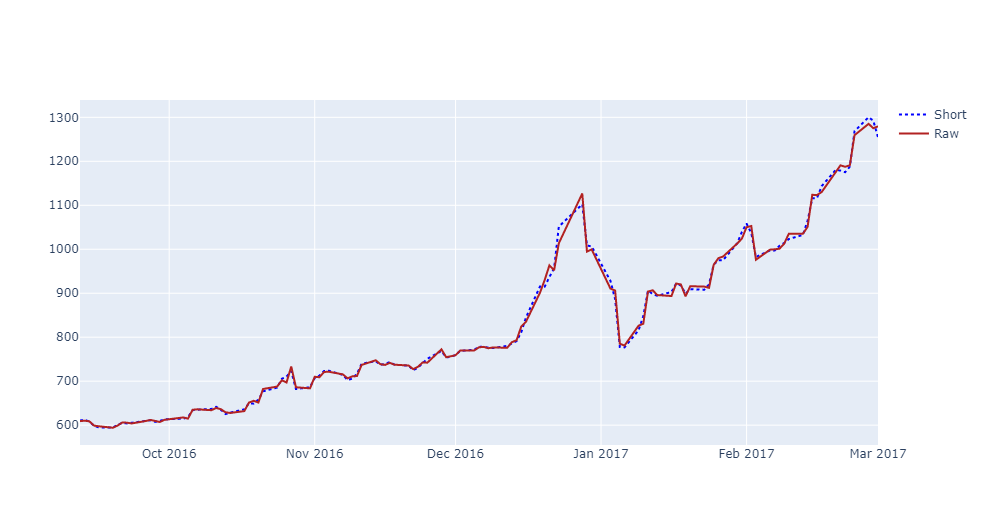
Towards a sliding window of five days in length, we conduct Stationary test: ?



**Fig 4 Fifth order sliding ADF window of Bitcoin and Gold**

The results of ADF test show the short term (ranging 5 days) appears fluctuation. Poly-regression is proved to be a powerful tool for short-term time series with high volatility and nonlinearity to forecast.

Here is part of the forecasting of Poly-regression randomly intercepted:



(a)



(b)

**Fig 6 Short-term polynomial predicted vs. actual values**

The figures 5 and 6 show that poly-regression performs well in fitting short-term fluctuation.

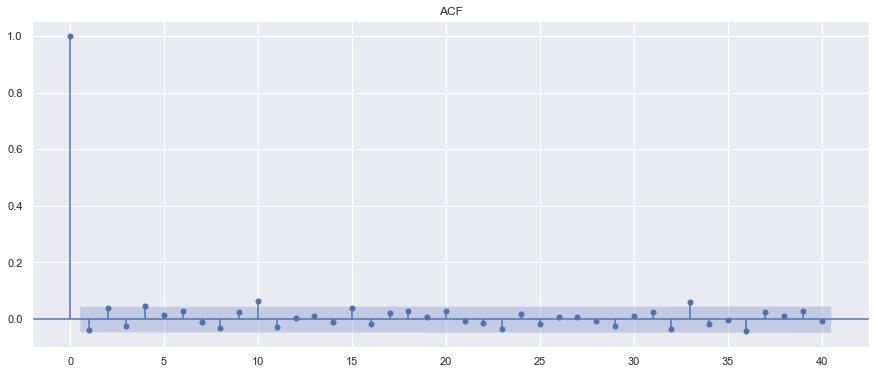
**5.2.2 Medium Term Income Factor by Fixed ARIMA**

The data series of the forecast object over time is considered as a random series, and a mathematical model is used to approximate this series. Once the model is identified, it can predict the future values from the past and present values of the time series.

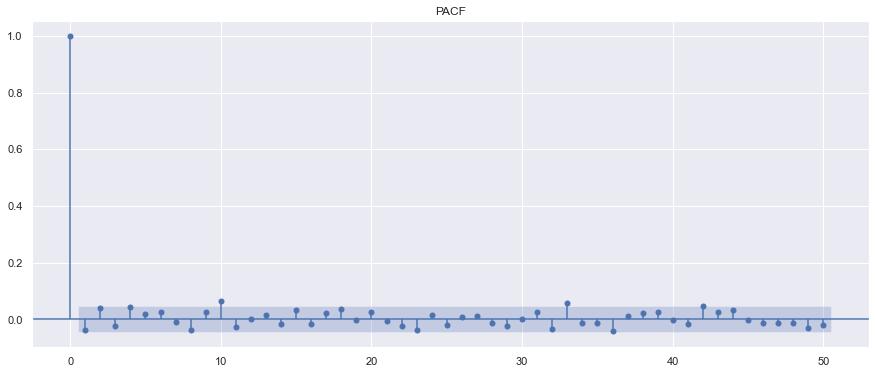
The ARIMA model contains three orders: , where: represents the number of lags (lags) of the time series data itself used in the forecasting model, represents the number of orders of differencing needed for the time-series data to be stable, and represents the number of lags of the prediction error used in the prediction model (lags). The formula of the ARIMA model is in Formula (8).

Usually, ARIMA needs smooth series fitted and a combination of ACF and PACF to determine the order. It is clear from Figure xx that the exponentially rising prices of bitcoin and gold are not stable, and this is corroborated by the ADF and KPSS in the data preprocessing section.

Obviously, the bitcoin price trend has a soundly different performance in the first half and the second half - a relatively slow rise in the first half and a dramatically volatile rise in the second half. Traditionally, ACF/PACF estimate the and values, but Fig shows a weak ARIMA relationship in the transformed data.



(a)



(b)

**Fig 7 (a) ACF of ARIMA (b) PACF of ARIMA**

Instead, in [] , ARIMA can be adopted for much shorter series and the change in order and has little effect on the error. ARIMA with fixed order as a structural function fitted by medium-period series is chosen to capture the trend.

**5.2.3 Long Term Income Factor via Bayesian Structural Time Series Model**

A structural time series model is defined by two equations including observation equation and latent state variables . They are defined:

The error terms and are Gaussian and independent of everything else. and are structural parameters.

In our model. , and are scalar value 1. The compromise, standing for the local level model is determined by

For a long term time series forecasting：

which .

Simply, an alternative is to replace the random walk with a stationary AR process based on the situation:

with,.

At this time, with the increase of the distance between the prediction time points, the distribution of tends to be:

hence can be expressed as:

is the trend component in long term, and will gradually revert to it.

Bayesian Structural time series models are flexible and modular. based on things like whether short or long term predictions are more important: whether the history price contains seasonal effects, and whether and how regressors are to be included. Benefit from the state identification and long length window, this part could control the risk.

**5.3 Risk Controlling through Conditional Value at Risk (*CVaR*)**

Value at risk (VaR)[] is defined as a given holding period and a given confidence level, the potential maximum loss that may be caused to a certain capital position, asset portfolio or institution when market risk factors such as interest rates and exchange rates change.

While VaR is a very useful tool, but it is not the most comprehensive solution to the problem of managing price risk. This is because VaR cannot solve the problem of judging how much to lose within the specified time, it just roughly states the maximum amount we can lose. The VaR model also assumes independence between consecutive trading days. That is to say, it is assumed that transactions that occur today will not have any effect on transactions tomorrow. This is a flawed assumption, because whether the market is mean-reverting or trending, the quantitative investments we make definitely have an impact on the nature of the market. This assumption also indirectly results in the inability of VaR to explain a series of continuous times.

The CVaR model[] overcomes the shortcomings of the VaR model to a certain extent. It not only considers the frequency of exceeding the VaR value, but also considers the conditional expectation of the loss that exceeds the VaR value, which effectively improves the problem of the VaR model in dealing with the rear tail phenomenon of the loss distribution.

If represents the -day returns then , where Conditional Value-at-Risk, expressed as a percentage of the portfolio value, is given by:

Therefore, in order to derive CVaR for any continuous probability density function of , we need to integrate over till -day VaR (i.e. quantile). Now, one can find that:

is the conditional value-at-risk CVaR in the normal linear VaR model for a random variable over -day horizon where denotes the standard normal density function and the quantile of the standard normal distribution.

补充结果

# 8. Modifications of our model

# 9. Sensitivity Analysis

# 10. Strengths and Weaknesses

10.1 Strengths

* The Adapting Markowitz Pair-Portfolio model established in our paper is based on the Modern Portfolio Theory and research results of Conditional Value at Risk, thus it is relatively rigorous.
* The results of multi-temporal factor income strategy based on different terms income factor matches poly regression, fixed ARIMA, Bayesian Structural Time Series Model, which indicates our model is reasonable and effective.
* The entire model both models profit forecasting and takes into account the presence of risk. It is able to steadily traverse bull and bear markets in the unpredictable financial trading market.

10.2 Weaknesses

* Ignoring the fluctuations during the day, we use closing price to describe t he market price of the day, which may reduce the accuracy of our model.
* We argue that all the market is purely rational, neglecting the public opinion, the orientation of politics, which may exert on our risk controlling strategy.

Memorandum(新的一页)

# References