|  |  |  |
| --- | --- | --- |
| **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:

图片

**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

# 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 |

# 4. Data Preprocessing

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

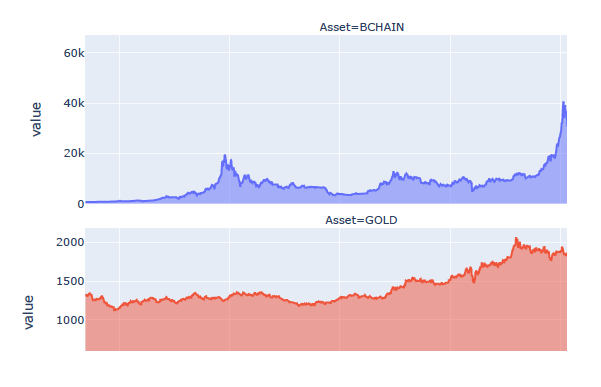
公式也要写（先不管）

**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 value on certain days. On account of abundant five-year data, we ignore the two days.

Subsequently, we visualize the modified data as follows:

****

**Fig 2**

**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 denote 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 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and augmented Dickey–Fuller test (ADF) tests.

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, respectively, 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  (after difference) | KPSS  (after 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 difference method and we present the results in table 2. Obviously, the outcomes satisfy Alternative hypothesis. Then we adopt the results after 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. For the reason that the VAR class assumes that the passed time series are stationary, we take advantage of data after difference. 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:

The variables of the form indicate that variable's value time periods earlier and are called the "*i*th lag" of . The variable *c* is a *k*-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. 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.

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

Next, 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**

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 取名**

|  |  |  |
| --- | --- | --- |
|  | 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.

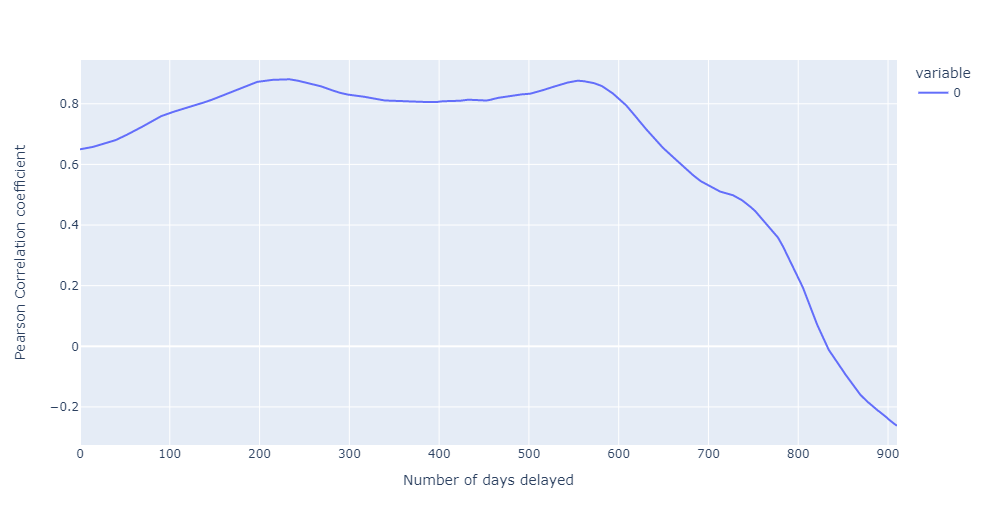
**4.3 Pearson Correlation Test**

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 show the Pearson correlation coefficients over time in Fig 3.

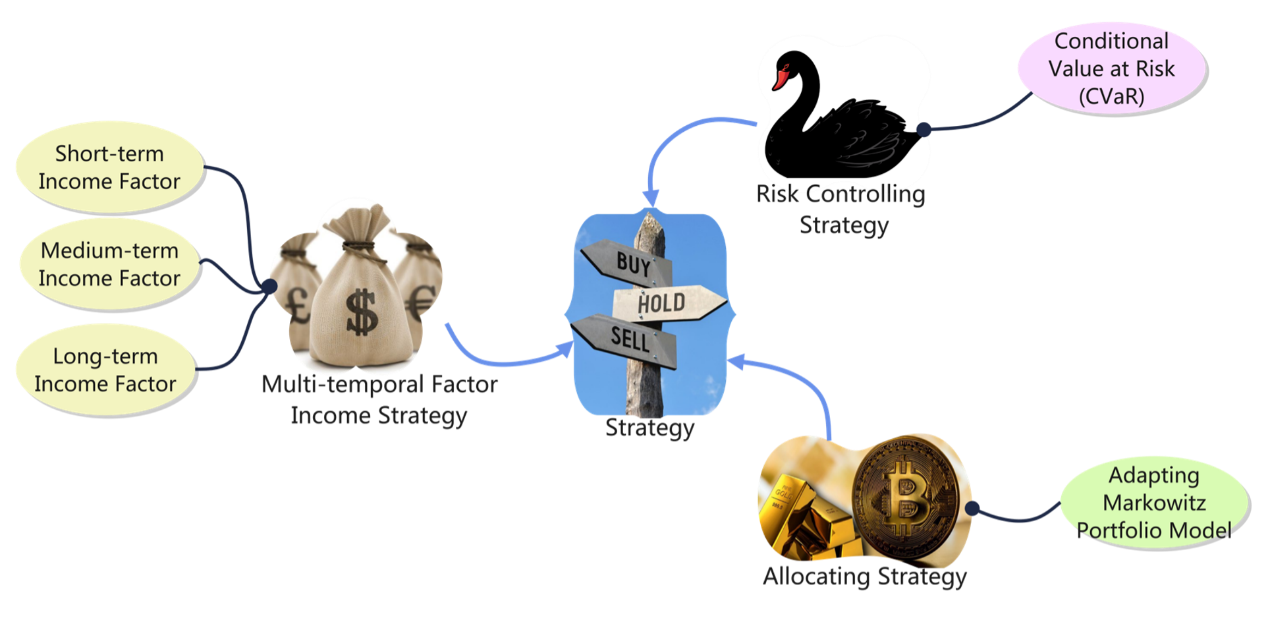


**Fig 3**（换图）

According to Fig 3, We can see that there was a delay of more than 200 days when Pearson's correlation coefficient reached its maximum. We 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 last such a long period of time.

# 5. Multi-temporal Factor Income Strategy

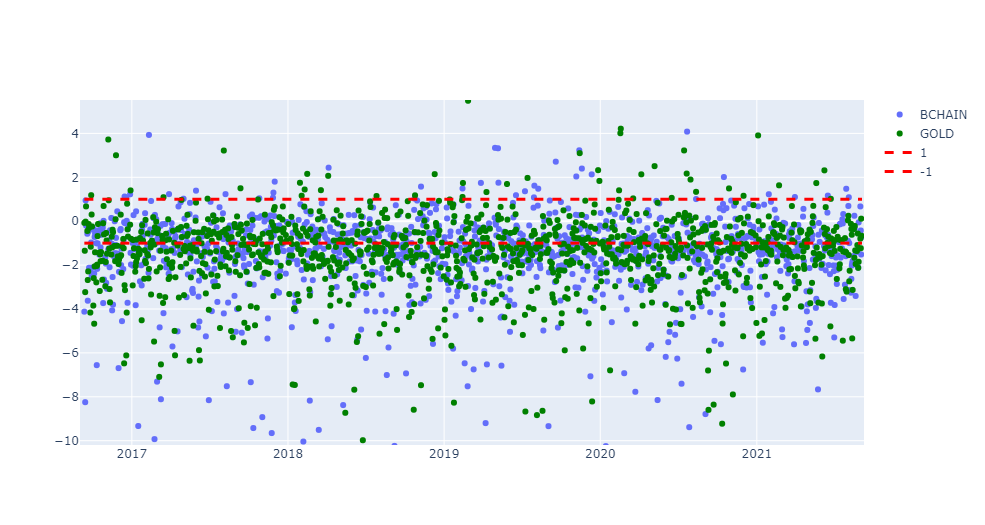
In most quant investment strategies, income model includes several factors to measure the product’s the ability of earnings like Fama-French Model. But the data available, history price merely, is limited compared to that required in Fama-French Model. Correspondingly, we estimate three factors based on time series of different lengths including short, medium, and long term income factors. Here come the details:



**Fig 4** 图片名称

**5.1 Short Term Income Factor based on Poly-regression**

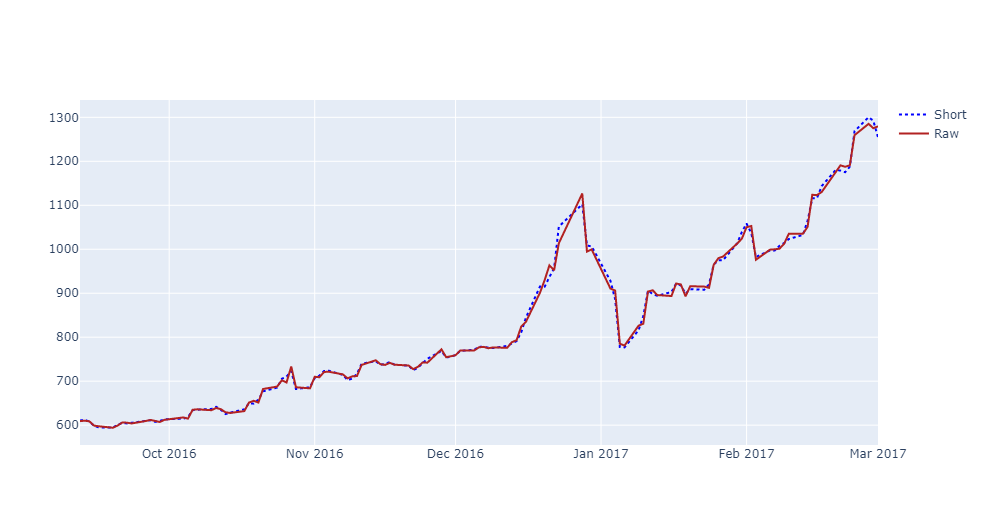
Towards a sliding window of five days in length, we conduct Stationary test:



**Fig 5** 图片名称

The results of ADF test shows 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:





# 8. Modifications of our model

# 9. Sensitivity Analysis

# 10. Strengths and Weaknesses

10.1 Strengths

10.2 Weaknesses

Memorandum(新的一页)

# References