| | | |
|---|---|---|
| **Problem Chosen**<br>C | **2022 MCM/ICM Summary Sheet** | **Team Control Number**<br>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 $\alpha\%$ of the amount traded and we assume that $\alpha_{gold} = 1\%$ and $\alpha_{bitcoin} = 2\%$. 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 |
| --- | --- |
| *ERI* | Economic recession index |
| *ESI* | Ecosystem sustainability index |
| *SHI* | Social habitability index |
| *FCI* | Fragility index based on the climate change |

| | |
|---|---|
| *CCI* | Climate change index |
| *C* | Pearson's contingency coefficient |
| *TC* | 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.

## 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:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t \#(1)$$

The variables of the form $y_{t-i}$ indicate that variable's value $i$ time periods earlier and are called the "$i$th lag" of $y_t$. The variable $c$ is a $k$-vector of constants serving as the intercept of the model. $A_i$is a time-invariant ($k \times k$)-matrix and et is a $k$-vector of error terms. The error terms must satisfy the following conditions: $E(e_t) = 0$. 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 $e_t$ is the residual given by $e_t = \rho e_{t-1} + \upsilon_t$, the Durbin-Watson statistic states that null hypothesis: $\rho = 0$, alternative hypothesis: $\rho \neq 0$, then the test statistic is

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})}{\sum_{t=2}^{T} e_t^2} \#(2)$$

where $T$ is the number of observations.

Since $d$ is approximately equal to 2 $(1 - \hat{\rho})$, where $\hat{\rho}$ is the sample autocorrelation of the residuals, $d$ = 2 indicates no autocorrelation. The value of $d$ always lies between 0 and 4. If the Durbin–Watson statistic is substantially less than 2, there is evidence of positive serial correlation. If $d$ > 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 $\rho_{X,Y}$ and its formula is as follows:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_{i=1}^{n}\left(x_i - E(X)\right)\left(y_i - E(Y)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - E(X)\right)^2} \cdot \sqrt{\sum_{i=1}^{n}\left(y_i - E(Y)\right)^2}} \#(3)$$

where $Cov(X,Y)$ represents covariance of X and B. $E(X)$ and $E(Y)$ indicate 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 = $5.5280117094571344 \times 10^{-219.}$

Next, we take time ductility of events into consideration and show the Pearson correlation coefficients over time in Fig 3.

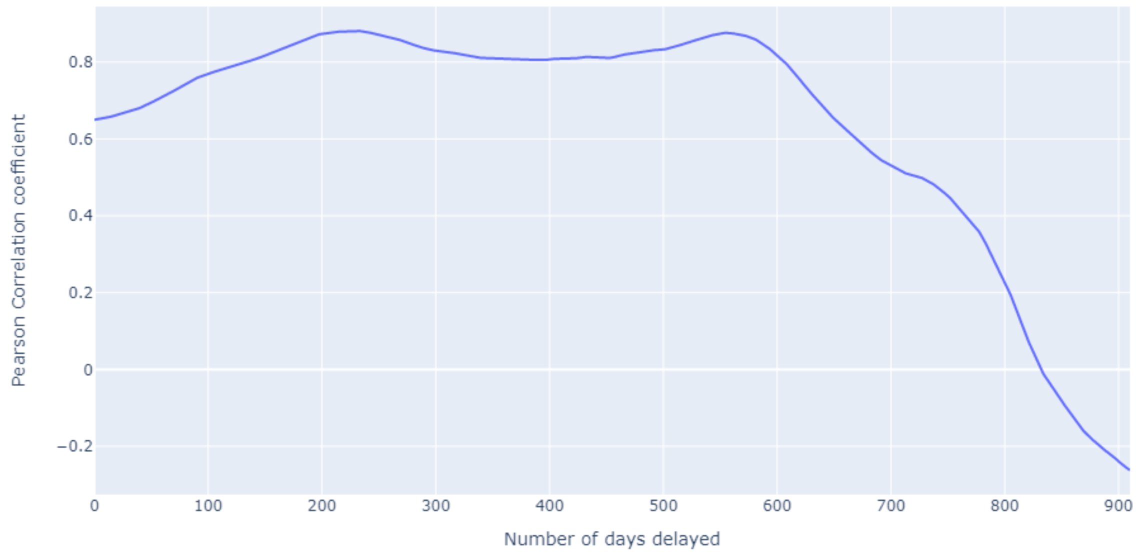


**Fig 3**（换图）

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:

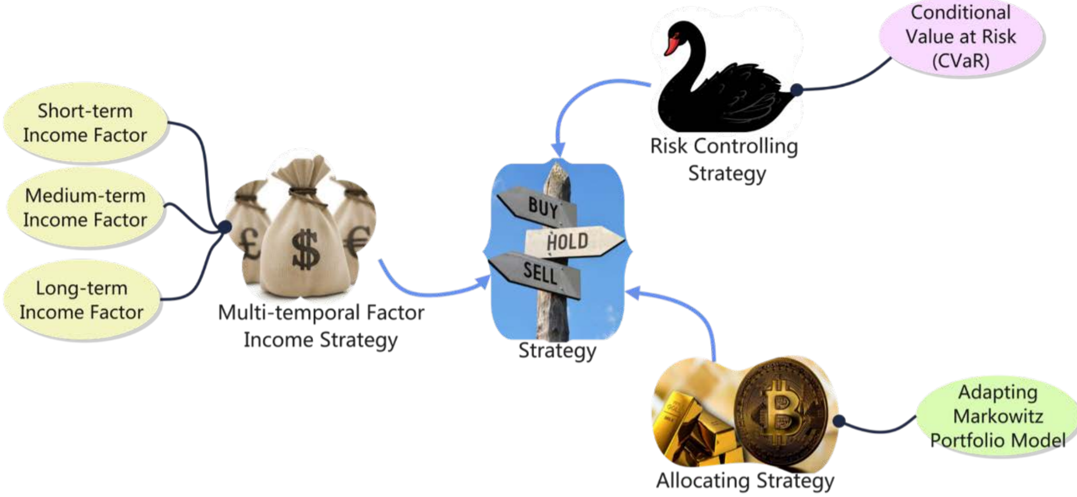


**Fig 4** 图片名称

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

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

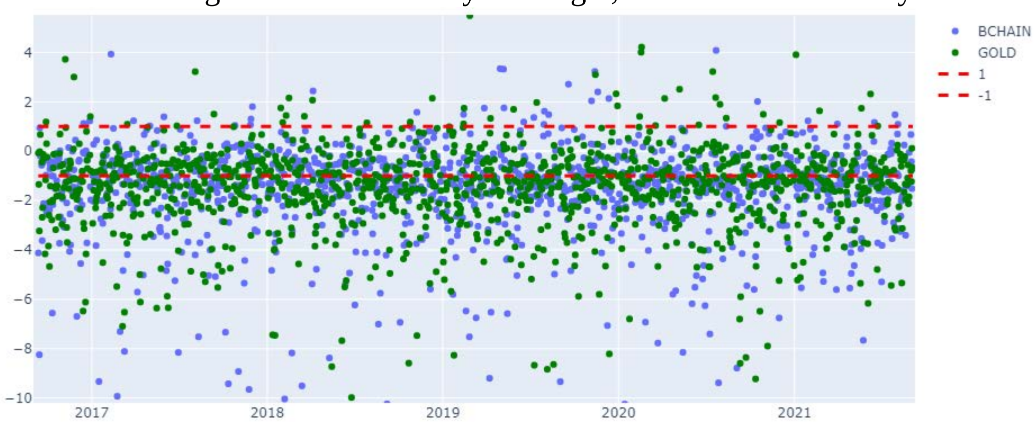


**Fig 5** 图片名称

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:


**Fig 6** 图片名称


**Fig 7** 图片名称

## 5.2 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.

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right) \#(4)$$

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

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 $p$ and $q$ values, but Fig shows a weak ARIMA relationship in the transformed data.

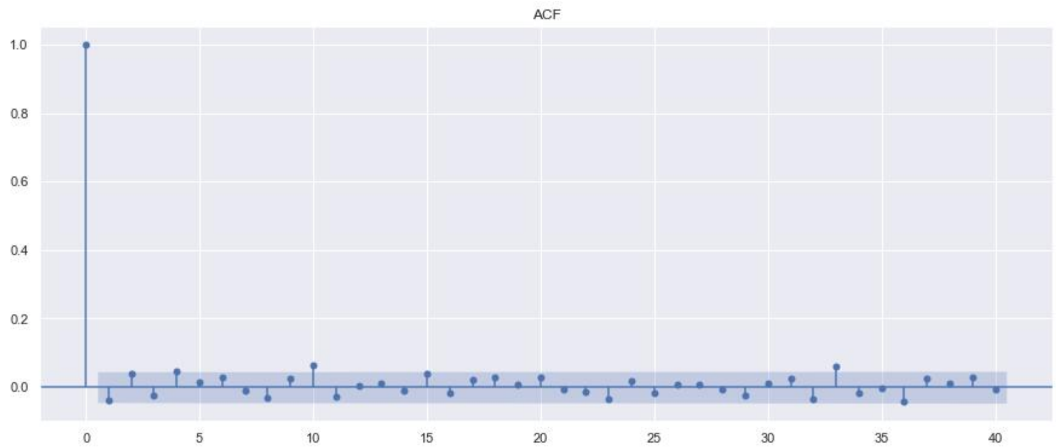


**Fig 8** 图片名称

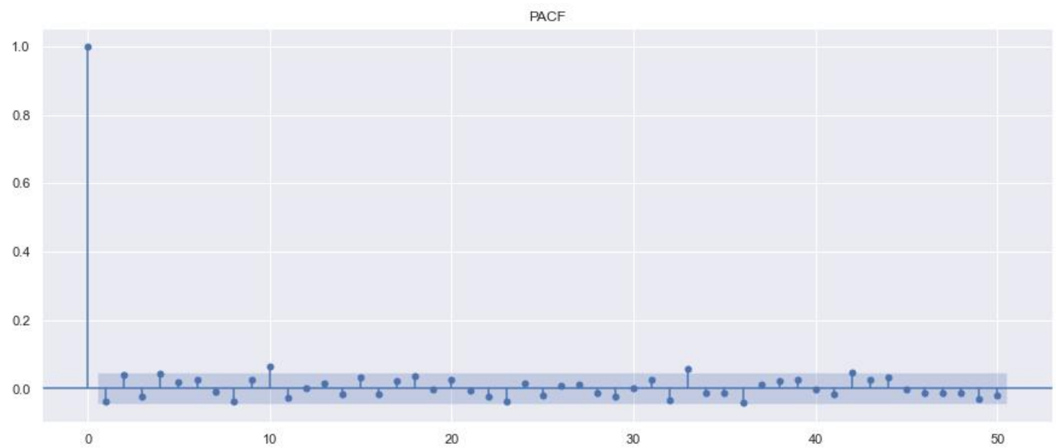


**Fig 9** 图片名称

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

### 5.3 Long Term Income Factor via Bayesian Structural Time Series Model

Finally, different in coming factors are summarized in Table 4:

**Table 4** 取名

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

## 6. Risk Controlling through Conditional Value at Risk (*CVaR*)

### 6.1 Brief Description

Generally speaking [], if an investment has shown stability over time, then the value at risk may be sufficient for risk management in a portfolio containing that investment. However, the less stable the investment, the greater the chance that VaR will not give a full picture of the risks, as it is indifferent to anything beyond its own threshold.

### 6.1.1 quedian

Conditional Value at Risk (CVaR) [] attempts to address the shortcomings of the VaR model, which is a statistical technique used to measure the level of financial risk within a firm or an investment portfolio over a specific time frame. While VaR represents a worst-case loss associated with a probability and a time horizon, CVaR is the expected loss if that worst-case threshold is ever crossed. CVaR, in other words, quantifies the expected losses that occur beyond the VaR breakpoint.

### 6.1.2 youdian

https://www.zhihu.com/search?type=content&q=CVaR

## 6.2 Model（调行间距）

If $X$ represents the $h$-day returns then $\mathrm{VaR}_{h,\alpha} = -x_{h,\alpha}$, where $P\left(X < x_{h,\alpha}\right) = \alpha$. Conditional Value-at-Risk, expressed as a percentage of the portfolio value, is given by:

$$\mathrm{CVaR}_{h,\alpha}(X) = -E\left(X \mid X < x_{h,\alpha}\right) = -\alpha^{-1} \int_{-\infty}^{x_{h,\alpha}} x f(x) dx \#(5)$$

Therefore, in order to derive CVaR for any continuous probability density function of $f(x)$, we need to integrate $xf(x)$ over $x$ till $100(1-\alpha)\%h$-day VaR (i.e. $x_{h,\alpha}$ quantile). Now, one can find that:

$$\mathrm{CVaR}_{h,\alpha}(X) = \alpha^{-1}\varphi\left(\Phi^{-1}(\alpha)\right)\sigma_h - \mu_h \#(6)$$

is the conditional value-at-risk CVaR in the normal linear VaR model for a random variable $X \sim N(\mu_h, \sigma_h^2)$ over $h$-day horizon where $\varphi(z)$ denotes the standard normal density function and $\Phi^{-1}(\alpha)$ the $\alpha$ quantile of the standard normal distribution.

## 6.3 Allocating Strategy from Adapting Markowitz Portfolio Model（调行间距）

Markowitz's modern asset allocation theory takes as input the expected return and covariance matrix of these investments and finds the optimal portfolio by optimizing the following objective function.

$$\max_{\omega_{eq}} \omega_{eq}{}'\mu - \frac{1}{2}\delta\omega_{eq}{}'\Sigma\omega_{eq}$$

where $\mu_0$ denotes the vector of expected returns of the investments, $\Sigma$denotes the covariance matrix of the investments, $\delta$ denotes the risk aversion coefficient of the investors, and $\omega_{eq}$ is the allocation weight of the investments in the portfolio.

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

$\omega^* = \Sigma^{-1}\mu$

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. 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. We refine the Markowitz model based on the expected return and risk calculated from the data we get:

define the exception and variance vectors of gold short, medium and long-term factors are:

$E_{SGOLD}, E_{MGOLD}, E_{LGOLD}, SD_{sGOLD}, SD_{MGOLD}, SD_{LGOLD}$, and $E_{SBitcoin}, E_{MBitcoin}, E_{LBitcoin}, SD_{SBitcoin}, SD_{MBitCoin}, SD_{LBitcoin}$

And $E_{Gold} = mean(E_{SGOLD}, E_{MGOLD}, E_{LGOLD})$,

$E_{Bitcoin} = mean(E_{SBitcoin}, E_{MBitcoin}, E_{LBitcoin})$;

$SD_{Gold} = \frac{sum(SD_{sGOLD},\ SD_{MGOLD},\ SD_{LGOLD})}{9}$,

$SD_{Bitcoin} = \frac{sum(SD_{SBitcoin},\ SD_{MBitcoin},\ SD_{LBitcoin})}{9}$,

so, we get $\mu = \begin{pmatrix} E_{Gold} \\ E_{Bitcoin} \end{pmatrix}$ and $\Sigma = \begin{pmatrix} SD_{Gold} & SD_{Bitcoin} \\ SD_{Bitcoin} & SD_{Gold} \end{pmatrix}$

# 8. Modifications of our model

# 9. Sensitivity Analysis

# 10. Strengths and Weaknesses

## 10.1 Strengths

-

## 10.2 Weaknesses

-

Memorandum（新的一页）

# References