

Inspect the Connection between Cryptocurrency and Stock through Markov Switching Process

HSIU-HSUAN, YEH

June 7, 2023

Abstract

The reason for this article choosing the markov switching assumption is that I am curious about if I can identify the direction of capital flow between the cryptocurrency and stock markets. This kind of question is always difficult since it's hard to decide which leads the others. However, we can still somehow have some idea about the relationship of these markets. The main results of this article is that we can observe the close relation in the early period which means the switching process of markets are highly overlapped, while in the recent year, the switching process starts to be isolated.

1 Introduction

Recently, cryptocurrency has gained a lot of attention. Some are optimistic to the development of the so-called decentralized finance due to the flexibility and the freedom of sharing information. However, in the other side of the "freedom" is the lack of regularity. Therefore, Some scandal and fraud has taken place, for example, the collapse of Terra Luna, bankruptcy of FTX which

is the previous second-largest cryptocurrency exchange before bankruptcy, to name just a few. It seems that there are large amounts of to speculate capital flows in the cryptocurrency market imposing the crypto assets in a high risk of price fluctuation. What's more, due to the tight connection in financial market, traditional assets will be influenced in some sense. For example, if crash happens in cryptocurrency market, there might be a large capital flowing to the other financial markets such as the stocks and bonds for the sake of avoiding risk.

The purpose of this research is to examine the connection between cryptocurrencies and stock prices using a Markov switching process. The reason that I assume there is Markov switching in both market is that there two common type of financial markets which is the so called "bear market" and "bull market", Hence, it seems reasonable to assume there are two hidden states for both markets. Furthermore, I also curious about if the ordered of changing process which the cryptocurrency markets leads the stock markets based on the prior knowledge discussed above can be observed.

The structure of this article is as follows: I first examine the characteristics of the return of S&P500 and BTC-USD such as stationarity condition, mean and standard error. Basic OLS of autoregressive is considered as well. Afterwards, I combine the autoregressive property and markov switching assumption to model the assets return.

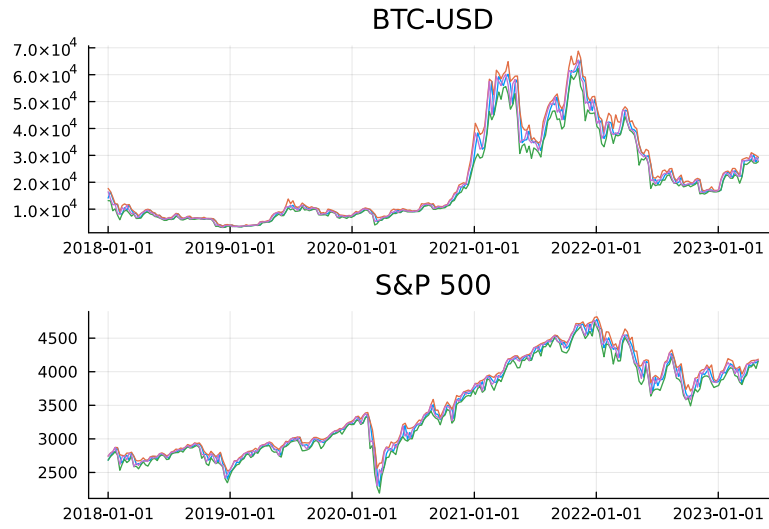


Figure 1: the historical weekly price of BTC-USD and S&P500

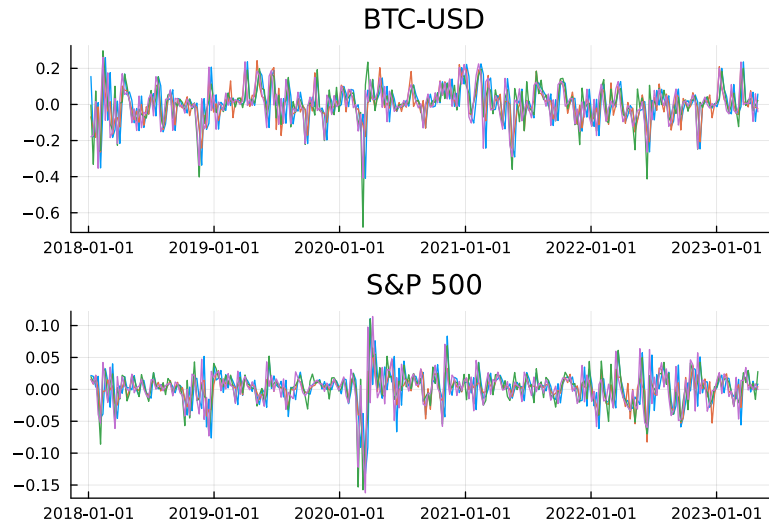


Figure 2: the historical weekly return of BTC-USD and S&P500

2 Data

Both the S&P500 and BTC-USD historical weekly price start from 2018/1/1 to 2023/5/1 is download from the website of the Yahoo Finance with 278 observations. Due to the non-stationary issues of the price, I transform the price data to log return. Demonstrate the plot in Figure 1. After transformation, the mean of the S&P500 weekly return is about 0.15047% and the BTC-USD's is about 0.19189%. Though BTC-USD seems to have higher return, the sample standard deviation differ quite a lot. The former is 0.02835 while the latter is 0.10143. It turns out that the BTC-USD is more volatile than the S&P500, in average and thus, suffering from the risk of high return fluctuation.

To have the basic understanding about return data of both S&P500 and BTC-USD I conduct the OLS by assuming the existence of auto correlation for both S&P500 and BTC-USD. In the following section I will use the notation of *spx* and *btc* to specify the weekly return of S&P500 and BTC-USD. A thousand miles begin with a single step, let's first fit AR model and later I will broaden the assumption to more general case which is the MS-AR model. To decide the proper lag for the AR model, inspecting the autocorrelation plot is always a good starting point. As shown in figure 2 and 3, we can see that the autocorrelation for *btc* persists much longer comparing to the *spx*. However, it's quite difficult to interpret such the long persistence. Therefore, I only consider lag period of 5 weeks for *spx* and 9 weeks, around 2 month for *btc* in order to fit the AR model.

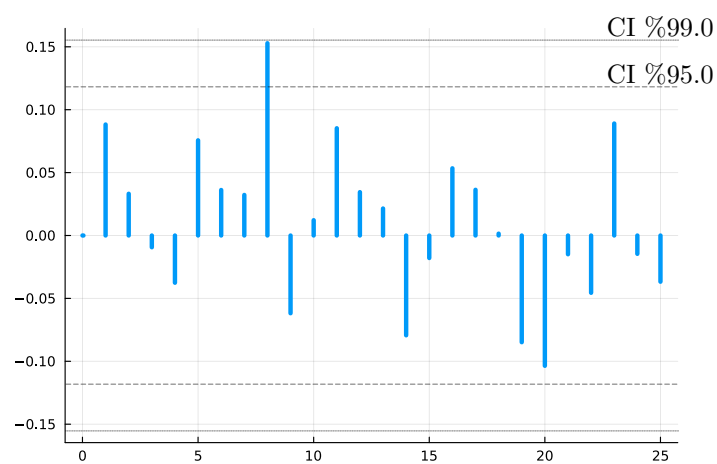


Figure 3: auto correlation function of the BTC-USD weekly return

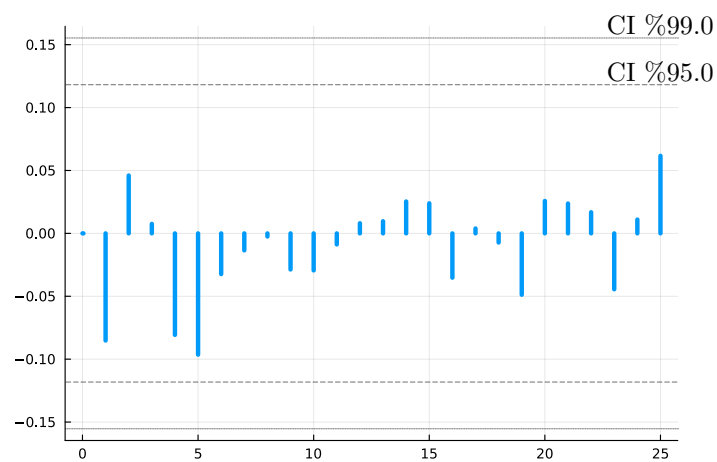


Figure 4: auto correlation function of the S&P500 weekly return

Before conducting the OLS, we need to ensure the stationarity condition holds. The hypothesis test I choose is the Augmented Dickey Fuller test with constant deterministic for both situation, lag amount of 5 and 9 period for spx and btc. The ADF statistic is -7.95658 and -4.60413. Because both reject the null hypothesis that there is uni-root, the stationarity requirement is satisfied. The following is the specification for the OLS. The reason that I consider the constant deterministic is that I first make the assumption, and then decide whether to lose it or not. After all, the 0 is just the special case for the constant deterministic.

$$spx_t = \phi_0 + \sum_{i=1}^p \phi_i spx_{t-i} + \epsilon_t$$

Variables	(1) AR(1)	(2) AR(4), $\beta_0 = 0$	(3) AR(4)	(4) AR(5), $\beta_0 = 0$	(5) AR(5)
constant	0.00158 (0.00170)		0.00165 (0.00172)		0.00205 (0.00171)
spx_{t-1}	-0.087531 (0.06006)	-0.07609 (0.06046)	-0.07894 (0.06055)	-0.09447 (0.06027)	-0.09868 (0.06033)
spx_{t-2}		0.04725 (0.0606)	0.04403 (0.0607)	0.05334 (0.06007)	0.04956 (0.06011)
spx_{t-3}		0.01386 (0.06058)	0.01054 (0.06069)	0.02231 (0.06008)	0.01852 (0.06012)
spx_{t-4}		-0.07834 (0.06037)	-0.08142 (0.06047)	-0.08272 (0.06)	-0.08671 (0.06005)
spx_{t-5}				-0.1086* (0.05998)	-0.11274* (0.06003)

Table 1: OLS for spx

In terms of AR(5) with zero mean, the t statistics for spx_{t-1} is -1.57, p-value is 0.1182 which is slightly insignificant in terms of the 10% significant level.

$$btc_t = \phi_0 + \sum_{i=1}^p \phi_i btc_{t-i} + \epsilon_t$$

Variables	(1) AR(1)	(2) AR(5), $\beta_0 = 0$	(3) AR(5)	(4) AR(9), $\beta_0 = 0$	(5) AR(9)
constant	0.00239 (0.00606)		0.00420 (0.006)		0.00308 (0.00589)
btc_{t-1}	0.08791 (0.05974)	0.08510 (0.06088)	0.08316 (0.061)	0.12168** (0.06156)	0.1208* (0.06167)
btc_{t-2}		0.01046 (0.05965)	0.01031 (0.05972)	0.01930 (0.06079)	0.01831 (0.06091)
btc_{t-3}		-0.03189 (0.05951)	-0.03523 (0.05958)	-0.05182 (0.06082)	-0.05277 (0.06093)
btc_{t-4}		-0.03249 (0.05903)	-0.04106 (0.05909)	0.01843 (0.06007)	0.01714 (0.0602)
btc_{t-5}		0.08378 (0.05906)	0.08203 (0.05913)	0.06693 (0.05975)	0.06568 (0.05988)
btc_{t-6}				0.01809 (0.05847)	0.01736 (0.05857)
btc_{t-7}				0.01323 (0.05848)	0.01242 (0.05854)
btc_{t-8}				0.14785** (0.05893)	0.1475** (0.05902)
btc_{t-9}				-0.08995 (0.05896)	-0.08998 (0.05904)

Table 2: OLS report for btc

3 Specification

Inherit from the previous OLS AR(p) specification, let's add the component of markov swithching process and name it as MS-AR(p). I assume there are two kinds of hidden state for each period t, and for different state, the coefficient differs as well. The variance is considered to be homogeneous for the reason of simplicity.

$$s_t \in \{0, 1\} \quad (1)$$

$$spx_t = (\mathbb{1}(s_t = 0)\phi_0^0 + \mathbb{1}(s_t = 1)\phi_0^1) + \sum_{i=1}^p \phi(s_{t-i})spx_{t-i} + \epsilon_t \quad (2)$$

$$= (1 - s_t)\phi_0^0 + s_t\phi_0^1 + \sum_{i=1}^p \phi_i(s_t)spx_{t-i} + \epsilon_t \quad (3)$$

$$\phi_i(s_t) = (1 - s_t)\phi_i^0 + s_t\phi_i^1 \quad (4)$$

$$btc_t = \phi_0(s_t) + \sum_{i=1}^p \phi_i(s_t)btc_{t-i} + \epsilon_t \quad (5)$$

The transition probability matrix:

$$P = \begin{bmatrix} \mathbb{P}(s_t = 0 | s_{t-1} = 0) & \mathbb{P}(s_t = 1 | s_{t-1} = 0) \\ \mathbb{P}(s_t = 0 | s_{t-1} = 1) & \mathbb{P}(s_t = 1 | s_{t-1} = 1) \end{bmatrix} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

$$\sum_{j=0}^1 p_{ij} = 1$$

My main purpose is to identify the markov-switching property. The standard AR assumption will have such judge that the results of the estimation is a kind of mixture leading to the imprecise estimation. However, it's quite tricky to estimate MS-AR model some assumptions on the initial prediction probability are required. I will address this in next section. I first assume ϵ_t follows normal distribution to apply Quasi-maximum likelihood estimation. I will examine the assumption through analysis on residual to see if it make sense.

4 Estimation

Denote: $S_T = \{s_T, \dots, s_{t+1}, s_t, s_{t-1}, \dots, s_1\} = \{0, 1\}^T$, is the hidden state of the whole period. $Y_T = \{y_T, \dots, y_{t+1}, y_t, y_{t-1}, \dots, y_1\}$ which can be the observed spx or btc of the whole period.

In the derivation step of quassi log-likelihood function, there are two sorts of probability:

1. prediction probability: $\mathbb{P}(s_t = i | Y_{t-1})$
2. filtering probability: $\mathbb{P}(s_t = i | Y_t)$

The parameters to be estimated: $\theta = (p_{00}, p_{11}, \sigma_e^2, \{\phi_i^s | s \in 0, 1\}_{i=0}^p)'$

$$L(\theta) = \frac{1}{T} \sum_{t=1}^T \log f(y_t | Y_{t-1}, s_t, \theta)$$

where f is the pdf of ϵ_t , and since s_t is known we need to take the expectation and get the quassi log-likelihood $L'(\theta) = \frac{1}{T} \sum_{t=1}^T \log f(y_t | Y_t - 1, \theta)$.

$$f(y_t | Y_t - 1, \theta) = \mathbb{P}(s_t = 0 | Y_{t-1}) \mathbb{P}(y_t | Y_{t-1}, s_t = 0; \theta) + \mathbb{P}(s_t = 1 | Y_{t-1}) \mathbb{P}(y_t | Y_{t-1}, s_t = 1; \theta)$$

$$\mathbb{P}(s_t = j | Y_{t-1}) = p_{0j} * \mathbb{P}(s_{t-1} = 0 | Y_{t-1}) + p_{1j} * \mathbb{P}(s_{t-1} = 1 | Y_{t-1})$$

$$\mathbb{P}(s_{t-1} = j | Y_{t-1}) = \frac{\mathbb{P}(y_{t-1} | s_{t-1} = j, Y_{t-2}; \theta) * \mathbb{P}(s_{t-1} = j | Y_{t-2})}{\mathbb{P}(y_{t-1} | Y_{t-2}; \theta)}$$

Through the backward induction, we need to consider the initial condition: $\mathbb{P}(s_1 = j | Y_0)$, and I follow suggestion from Hamilton (1994) set is as the limiting unconditional counterpart: the third column of $(A' A)^{-1} A' A = \begin{bmatrix} I_2 - P \\ 1_2' \end{bmatrix}$ where I is the identity matrix and P is the transition matrix, $1_2 = [1, 1]'$

Tips for coding the quassi log-likelihood function is to set the initial condition first, and follow the equation mentioned above sequentially.

After finish the construction of the quassi log-likelihood function, I use the BFGS to optimize the quassi log-likelihood. Moreover, the parameters p_{00}, p_{11} should in the $[0, 1]$ interval and σ_ϵ^2 should greater or equal to 0 so I take advantage of the normal CDF and exponential function to meet the restrictions. Speaking of asymptotic standard error, I first calculate the hessian matrix of the quassi log-likelihood function evaluating it at MLE, and then apply the delta method to get the correct standard error of $p_{00}, p_{11} \cdot \sigma_\epsilon^2$

when it comes to model selection, AIC and BIC are two types of common criterion.

$$AIC = T \log\left(\frac{SSR}{T}\right) + 2(2(p+1) + 3) \quad (6)$$

$$BIC = T \log\left(\frac{SSR}{T}\right) + (2(p+1) + 3) \log T \quad (7)$$

where T is the number of observations, SSR is the sum of squared residuals and $(2(p+1)+3)$ is the number of estimated parameters. To calculate the sum of squared residuals, the $\mathbb{1}(s_t = i)$ should be estimated. My personal selection is using smoothing probability which is specified by Kim(1994) as the estimator.

$$\hat{\mathbb{1}}(s_t = 0) = \mathbb{P}(s_t = 0|Y_T) \quad (8)$$

$$\hat{\mathbb{1}}(s_t = 1) = \mathbb{P}(s_t = 1|Y_T) \quad (9)$$

The derivation of smoothing probability:

$$\mathbb{P}(s_t = i|Y_T) = \mathbb{P}(s_{t+1} = 0|Y_T) \mathbb{P}(s_t = i|s_{t+1} = 0, Y_T) + \mathbb{P}(s_{t+1} = 1|Y_T) \mathbb{P}(s_t = i|s_{t+1} = 1, Y_T)$$

$$\begin{aligned} \mathbb{P}(s_t = i|s_{t+1} = j, Y_T) &= \frac{\mathbb{P}(s_{t+1} = j|s_t = i, Y_t) \mathbb{P}(s_t = i|Y_t)}{\mathbb{P}(s_{t+1} = j|Y_t)} = \frac{p_{ij} \mathbb{P}(s_t = i|Y_t)}{\mathbb{P}(s_{t+1} = j|Y_t)} \\ \Rightarrow \mathbb{P}(s_t = i|Y_T) &= \mathbb{P}(s_t = i|Y_t) \left(\frac{p_{i0} \mathbb{P}(s_{t+1} = 0|Y_T)}{\mathbb{P}(s_{t+1} = 0|Y_t)} + \frac{p_{i1} \mathbb{P}(s_{t+1} = 1|Y_T)}{\mathbb{P}(s_{t+1} = 1|Y_t)} \right) \end{aligned}$$

For coding part, I first get the prediction and filtering probability through the quassi maximum likelihood estimator, and the use the filtering probability at T as the initial point, calculate the remained part backward.

Variablues	(1) AR(1)	(2) AR(4)	(3) AR(5)
$(1 - s_t)\text{constant}$	0.00667(0.03181)	0.00868(0.03273)	0.00887(0.03098)
$(s_t\text{constant}$	-0.02175(0.12128)	-0.01583(0.06318)	-0.01509(0.05955)
$(1 - s_t)\text{spx}_{t-1}$	-0.28243(1.12692)	-0.34289(1.12399)	-0.34124(1.02077)
$s_t\text{spx}_{t-1}$	1.03979(4.09697)	0.7952(2.14944)	0.77522(1.99865)
$(1 - s_t)\text{spx}_{t-2}$		-0.02216(1.04064)	-0.03667(1.14148)
$s_t\text{spx}_{t-2}$		0.39114(1.73105)	0.38122(1.56481)
$(1 - s_t)\text{spx}_{t-3}$		-0.13722(1.10954)	-0.11505(1.11394)
$s_t\text{spx}_{t-3}$		0.26273(1.58801)	0.28116(1.55246)
$(1 - s_t)\text{spx}_{t-4}$		-0.22786(1.15642)	-0.26006(1.02341)
$s_t\text{spx}_{t-4}$		0.309134(1.74543)	0.30733(1.87587)
$(1 - s_t)\text{spx}_{t-5}$			-0.13458(1.02849)
$s_t\text{spx}_{t-5}$			-0.01249(2.17433)
p_{00}	0.82308(1.57008)	0.63659(1.99715)	0.65189(1.87251)
p_{11}	0.12336(1.62135)	0.18717(1.51741)	0.2035(1.47203)
σ_ϵ^2	0.00052(0.00087)	0.00043(0.00072)	0.000421(0.0007)
AIC	-2139.36473	-2165.29219	-2150.59379
BIC	-2113.9966	-2118.32152	-2096.45171

Table 3: normal quassi MLE report for spx

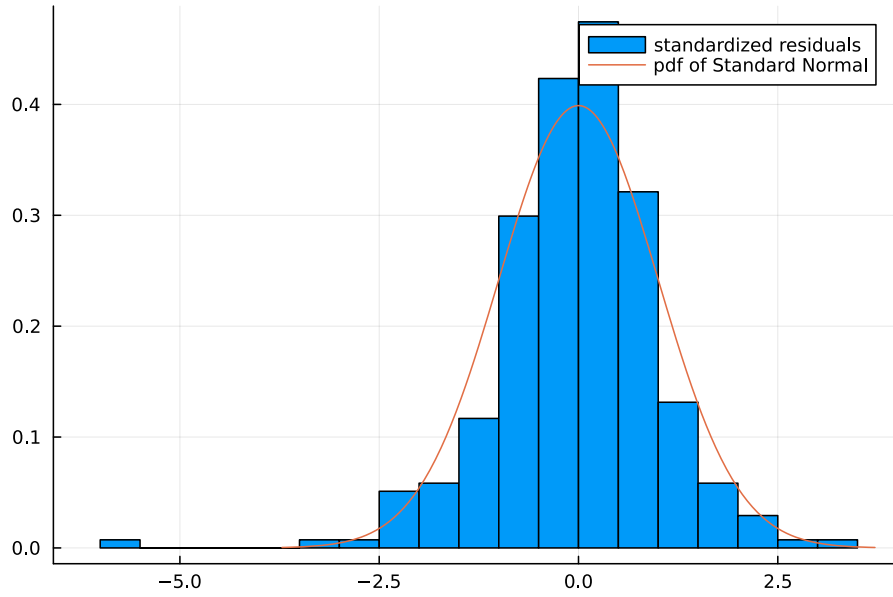


Figure 5: AR(4) residuals of spx

Because AR(4) has the lowest AIC and BIC, I choose it to do the residuals analysis to see whether the normal assumption is proper. To test whether the residuals follow a normal distribution, I decide to use the Jarque-Bera test and the test statistics is 180.239 which the null hypothesis is rejected. Though the null hypothesis has been rejected, I still think the normal distribution is not a bad approximation. In other word, with high possibility that the approximation won't deviate a lot changing the distribution assumption to student-t.

Variablues	(1) AR(1)	(2) AR(5)	(3) AR(9)
$(1 - s_t)\text{constant}$	0.00239(0.52578)	-0.0851(0.46047)	0.03227(0.201301)
$s_t\text{constant}$	0.00239(0.52578)	0.01678(0.11573)	-0.01984(0.2149)
$(1 - s_t)\text{btc}_{t-1}$	0.08791(3.04313)	0.68697(3.36902)	0.05143(1.5145)
$s_t\text{btc}_{t-1}$	0.08791(3.04313)	-0.01835(1.07007)	0.09761(1.91753)
$(1 - s_t)\text{btc}_{t-2}$		0.68697(5.44482)	-0.00111(1.60678)
$s_t\text{btc}_{t-2}$		-0.07558(1.25567)	0.04457(1.42198)
$(1 - s_t)\text{btc}_{t-3}$		-0.27521(2.69772)	-0.02662(1.73871)
$s_t\text{btc}_{t-3}$		-0.00082(1.07955)	-0.06516(1.67806)
$(1 - s_t)\text{btc}_{t-4}$		-0.11291(3.5996)	-0.03589(1.56324)
$s_t\text{btc}_{t-4}$		-0.03358(1.07279)	0.07499(1.67469)
$(1 - s_t)\text{btc}_{t-5}$		0.27714(5.59203)	0.15949(1.73148)
$s_t\text{btc}_{t-5}$		0.05071(1.32777)	0.00454(1.45309)
$(1 - s_t)\text{btc}_{t-6}$			0.23918(1.78003)
$s_t\text{btc}_{t-6}$			-0.20771(1.62786)
$(1 - s_t)\text{btc}_{t-7}$			0.10729(1.69226)
$s_t\text{btc}_{t-7}$			-0.07181(1.35164)
$(1 - s_t)\text{btc}_{t-8}$			0.15749(1.88957)
$s_t\text{btc}_{t-8}$			0.10623(1.32402)
$(1 - s_t)\text{btc}_{t-9}$			0.25688(2.27741)
$s_t\text{btc}_{t-9}$			-0.37145(2.17131)

Table 4: normal quassi MLE report for btc

Variablues	(1) AR(1)	(2) AR(5)	(3) AR(9)
p_{00}	0.5(2470)	0.05078(3.05193)	0.57067(2.15313)
p_{11}	0.5(2470)	0.83393(1.44172)	0.65415(3.61299)
σ_ϵ^2	0.01009(0.01427)	0.00685(0.01384)	0.00645(0.0122)
AIC	-1259.17703	-1378.71757	-1369.0413
BIC	-1233.80891	-1324.5755	-1286.36294

Table 5: normal quassi MLE report for btc

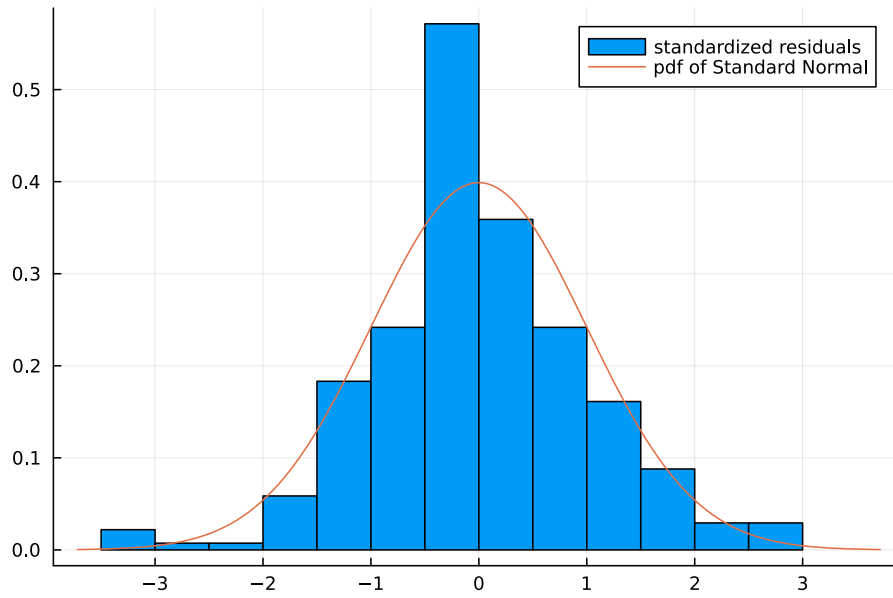


Figure 6: AR(5) residuals of btc

I conduct the Jarque-Bera test on the AR(5) of btx as well, the null hypothesis is rejected. In this case, student-t distribution may be a better choice. However, since the coefficient are all significant and the problem might stems from the small size of dataset, I didn't give it a shot.

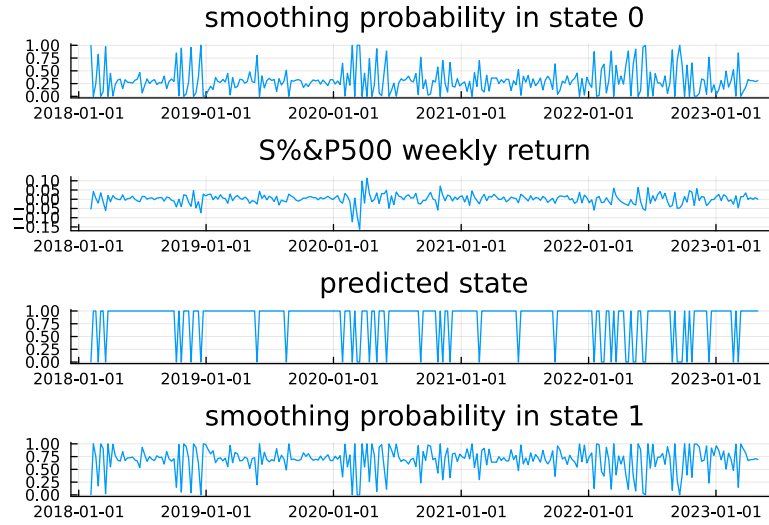


Figure 7: smoothing probabilitiy of spx modeled by AR(4)

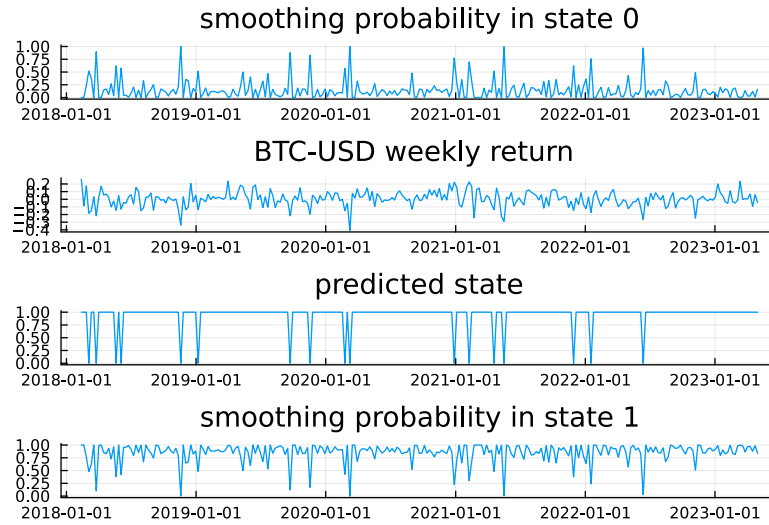


Figure 8: smoothing probabilitiy of btc modeled by AR(5)

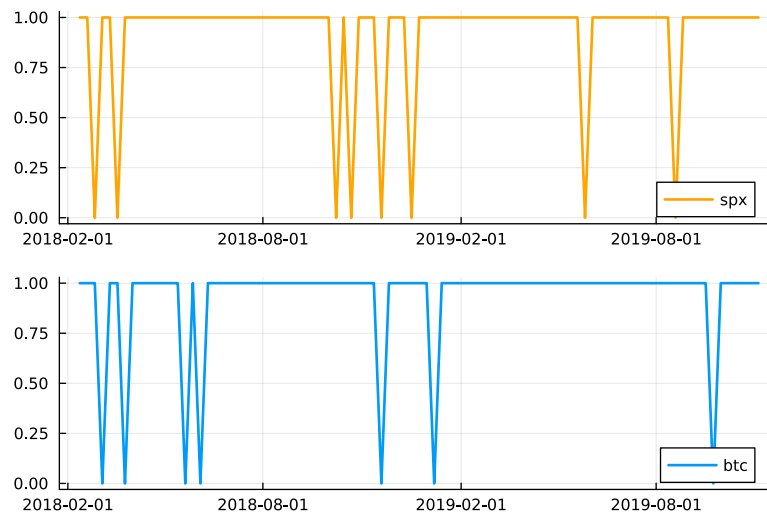


Figure 9: predicted state comparison(2018/02/12 - 2019-11-04)

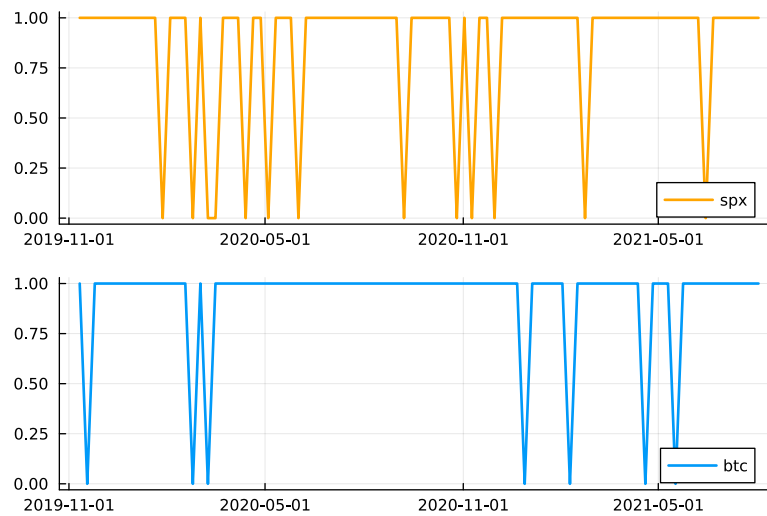


Figure 10: predicted state comparison(2019/11/11 - 2021-08-02)

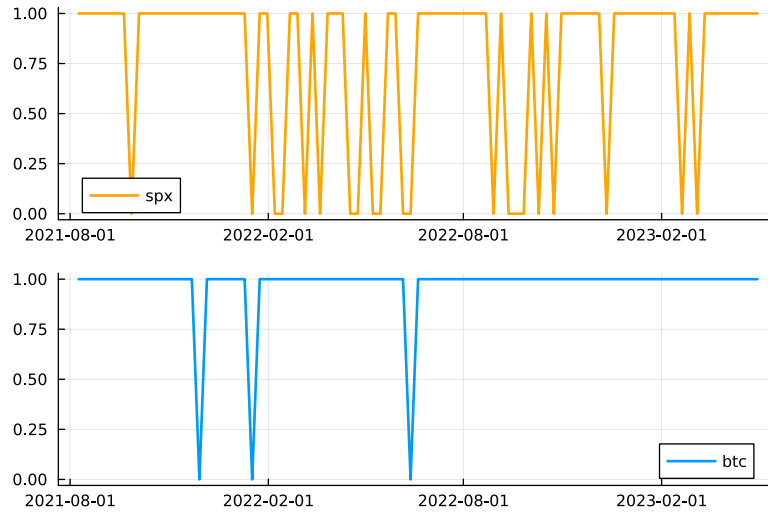


Figure 11: predicted state comparison(2021/08/09 - 2023-05-01)

5 Conclusion

After putting a lot of effort isolating the hidden states, we can observe that the relationship of changing process in two markets differ quite a lot to the past. Although technically, I get different state's smoothing probability in each markets, there is no actual meaning of "state 0" and "state 1". In fact the "state 0" varies a lot between two markets. Hence, I try to find have some insight and decide to reverse the state 0 as state 1 in stock markets. That is the pattern we observed from figure 9 to 11.

References

- [1] Hamilton (1994) Time Series Analysis
- [2] Kim (1994) Dynamic linear models with Markov-switching