A fractionally cointegrated VAR analysis of price discovery in Bitcoin spot and futures markets

Jinghong Wu

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science, Honours
in the Department of Economics
University of Victoria

Supervised by Dr. Ke Xu

For

Dr. Elisabeth Gugl, Honours Co-Advisor
Dr, Graham Voss, Honours Co-Advisor

Abstract

This paper adopts the Fractional Cointegrated Vector Autoregressive model to examine the price discovery of Bitcoin spot and futures prices from 18 December 2017 to 29 September 2019. The method compares the orthogonal component of the adjustment speed to the long-run equilibrium for both the spot price and futures price to test which market has more contribution to or takes the leadership in the price discovery process. In contrast to many other pieces of research in the price discovery of the Bitcoin market, the result of this paper shows the dominance of price discovery in the spot market. This may be caused by more transaction volume occurring in the Bitcoin spot market, which accords with findings of Baur and Dimpfl (2018), and Brandvold, Molnár, Vagstad, and Valstad (2015). To further explore the relationship between prices in Bitcoin spot and futures markets, I am going to examine the fractional cointegration between the two prices. Moreover, I will compare the FCVAR model to the traditional CVAR model to see which model better fits data in those two markets.

Keywords: Bitcoin, spot and futures markets, price discovery, fractional cointegration

Table of Contents

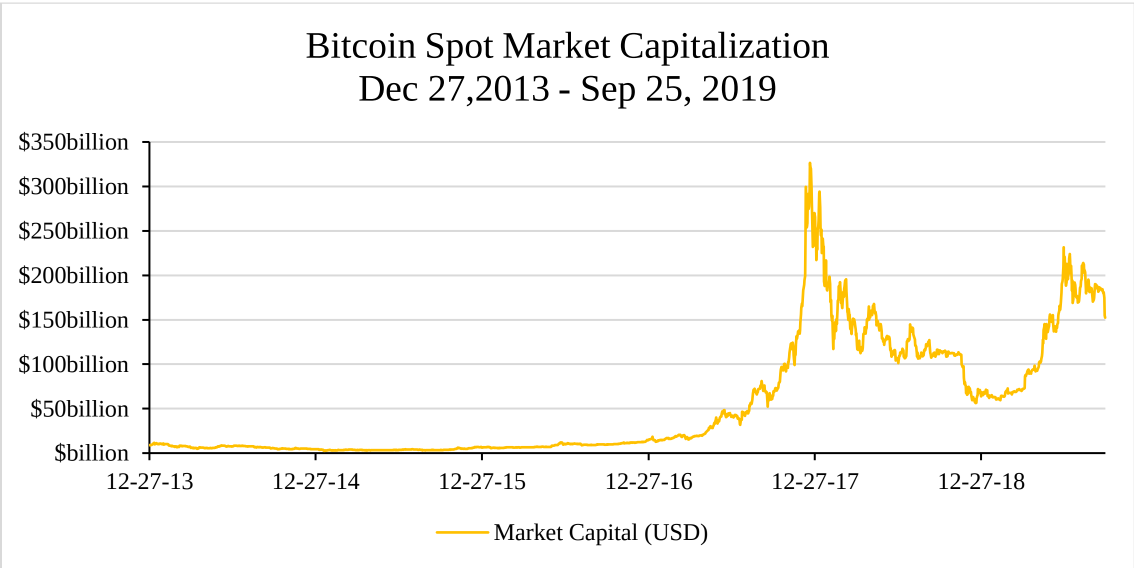
1. Introduction

I estimate the price discovery process for Bitcoin in the spot and futures markets, based on Gonzalo and Granger's (1995) permanent – transitory (PT) decomposition model and Hasbrouck's (1995) information share model. In this paper, the objective is to identify Bitcoin's fundamental or permanent value and determine which of the spot and futures prices provides more information for the adjustment of transitory prices to the fundamental value.

1.1 Motivation

Cryptocurrencies such as Bitcoin and Facebook's proposed Libra, continue to attract great attention. In recent years, governments and central banks have developed and promoted research on the operations of cryptocurrencies. See for example research at the Bank of Canada by Chiu and Koeppl (2019); see also the discussion in Livni (2019). With the advantage of low transaction cost, speediness, and block-chain technology for security, cryptocurrencies have gradually become promising future global payment tools and existing cryptocurrencies are heavily traded on well-established markets. According to Brauneis and Mestel's (2018) research among 73 cryptocurrencies, Bitcoin is the most unpredictable and thus most efficient cryptocurrency based on the random walk hypothesis. This paper is going to focus on Bitcoin and examine the spot and futures markets of this leading cryptocurrency. Figure 1 reports the spot market capitalization value for Bitcoin since 2013 and demonstrates both the volatility in this market as well as the increase in market size. It reveals the growing interest of investors as well as the potential high risk in the market. The study of price discovery of Bitcoin can provide useful information for more accurate Bitcoin price forecasts, from which market agents can generate more efficient investment and speculative trading strategies (Kavussanos and Nomikos (2003)). The information can also smooth supply and demand in the market and further lead to a more efficient and accurate pricing (Kavussanos and Nomikos (2003)).

Figure 1



Notes: Series are daily data retrieved from https://coinmarketcap.com. Data vintage is September 2019.

1.2 Price discovery approach

Using Gonzalo and Granger's (1995) PT decomposition, an observed price for an asset can be decomposed into two parts, the permanent component and the transitory component. The permanent component is an unchanged "common implicit efficient price" (Hasbrouck (1995)) which remains stable across different markets for an asset (Bitcoin in this paper). The transitory component can be seen as the result of "information share associated with a particular market" (Hasbrouck, 1995) that leads to variation in actual transaction prices. Unlike observed prices, price discovery is "the process of revealing an asset's permanent or fundamental value" (Dolatabadi, Nielsen, and Xu (2016)). It refers to the orthogonal component of the adjustment speed to the long-run equilibrium of an asset price. In the process of estimating the permanent price, I can use the (FCVAR) model to determine which of the spot and futures price contributes more to the behavior of the permanent price. This contribution is reflected in the orthogonal component of the speed of adjustment parameter in the Gonzalo and Granger approach.

The long-run equilibrium is specified here as the stationary cointegrated relationship of the spot and futures prices of Bitcoin. In the theory of spot-future parity, the cointegrated relationship of spot and futures prices is originated from the cost of carry. The futures price of the Bitcoin should be adjusted based on the spot price, by the cost of carry (including the cost of storing a commodity and other financial costs), the cost of money (investing the money to other places to generate interests), and dividends, so that there won't be any opportunity of arbitrage (making risk-free profits by just buying and selling the asset).

As price discovery is the process (an orthogonal component of the adjustment speed) of finding the fundamental value of an asset, when the asset is traded in multiple markets, different markets involved in the process will contribute differently to price discovery. The reason can be different market regulation, information disclosure, transaction cost, and speed, etc. We denote the market which "moves first" (Kapar and Olmo (2019)) or whose "product moves closer" (Kapar and Olmo (2019)) to the fundamental value as taking the leadership or dominating in the price discovery.

1.3 Estimation with FCVAR instead of CVAR

This paper applies Johansen's (2008) fractionally cointegrated vector autoregressive (FCVAR) model to analyze the price discovery of Bitcoin spot and futures markets. When talking about FCVAR, it is inevitable to introduce the more widely studied and well-known non-fractional cointegrated vector autoregressive (CVAR) model (Johansen (1995)). The spot price series and the futures price series are both non-stationary and integrated of degree 1, denoted $I(1)$. In theory, taking the first difference will make them both become stationary, which is $I(0)$. The spot price and futures price are cointegrated if the linear combination of the two non-stationary series is stationary or $I(0)$. That is the basic idea of the CVAR model. A CVAR model only uses integers in the integration degree parameter, while an FCVAR model allows fractional values for the parameter and any positive real number for the change of integration degree. So, the FCVAR model is more general and provides more possibilities when analyzing the relationships between different variables. It's "a natural methodology for the analysis of multiple fractional time series variables by generalizing the CVAR model to fractional time series process" (Dolatabadi et al (2016)).

1.4 Literature Review

Existing papers employ different methods for price discovery on Bitcoin. For instance, Brandvold, Molnár, Vagstad, and Valstad (2015) investigate the influence of trading markets on the price discovery of Bitcoin. They use the method of de Jong et al (2001) to investigate the price discovery process among seven exchanges (including Bitstamp, BTC-e (Btce), Mt.Gox (Mtgox), and another four popular bitcoin exchanges in the world). In the study, they find that Mtgox and Btce are dominating in the price discovery process. Instead of investigating the price of Bitcoin among different trading markets, as the future contract of Bitcoin launched in 2017, more and more researchers turn to the study of price discovery in the spot and futures market. Kapar and Olmo (2019) use the Hasbrouck (1995) information share model and Gonzalo and Granger's (1995) common factor component to test the price discovery leadership between the spot and futures market of Bitcoin. After calculating the information share in Hasbrouck (1995) and Gonzalo and Granger's (1995) methods, they conclude Bitcoin futures price takes the leadership in the price discovery process. Baur and Dimpfl (2018) do similar research in Bitcoin spot and futures markets using the CVAR model. Their conclusion is contradictory to the results of Kapar and Olmo's

(2019), with the spot market-leading in price discovery. In the literature, price discovery leadership usually happens in the futures market because of the lower transaction cost, fast speed, and more liquidity. For example, the study of 5 commodities of Figuerola-Ferretti and Gonzalo (2010) and the oil market of Foster (1996) shows futures dominance in price discovery. However, spot prices can sometimes take the leadership in price discovery if it is more active and has larger trading volume, such as the study of foreign exchanges of Chen and Gau (2010). In this paper, the price discovery coefficient in my model shows the spot market dominates the price discovery process of Bitcoin.

I use the method of fractionally cointegrated vector autoregressive (FCVAR) model for this paper to study the price discovery in the Bitcoin spot and futures market. Johansen (2008) adopts the fractional lag operator in mathematics and defines the fractional process in the vector autoregressive (VAR) model, which becomes the current FCVAR model. The FCVAR model is advantageous with the ability to adopt fractional values to parameters, compared to which, the traditional CVAR model could only adopt integers for the integration order parameters. The inference on coefficients in the FCVAR model is proved consistent with the maximum likelihood estimators (Johansen and Nielsen (2012)). Dolatabadi, Nielsen, and Xu (2016) prove the superiority of the FCVAR model over the CVAR model in the research of price discovery in spot and futures markets for five different commodities. They find high significance in the fractional integration parameter, which indicates the necessity to consider the fractionality in the formulation. Besides, in their research, fewer lags are required in the FCVAR model, so it can better avoid possible large discrepancies of the large lag length. All those qualities reflect the usefulness of the FCVAR model[1].

[1] Nielsen and Popiel (2016) provide a MATLAB program for the FCVAR model test statistics and calculation. My research in this paper is going to use this program.

2. Methodology

2.1 The Equilibrium Conditions

I am going to use the fractional cointegration equilibrium model of Dolatabadi, Nielsen, and Xu (2016), a variation based on the equilibrium model of spot and futures prices proposed by Garbade and Silber (1983) and Figuerola-Ferretti and Gonzalo (2010).

In accordance with the spot-futures parity condition, first, there is the model of equilibrium prices with infinitely elastic arbitrage. It means that under perfect market conditions, except regular financing costs, there are no arbitrage costs, such as taxes and restriction on short (borrowing) sales. For more details of the assumption, please see Garbade and Silber (1983) and Figuerola-Ferretti and Gonzalo (2010). The equilibrium model is:

$$f_t = s_t + r_t + c_t, \tag{1}$$

where f_t is the log-futures price, s_t is the log-spot price, r_t denotes the continuously compound interest rate, and c_t denotes the cost of carry. All variables measured at time t.

It is assumed that (1) $r_t = \bar{r} + u_{rt}$, where $\bar{r}$ is the mean of the continuously compound interest rate and u_{rt} is a stationary ($I(0)$) disturbance term with mean zero and finite positive variance; (2) $c_t = \bar{c} + u_{ct}$, where $\bar{c}$ is the mean of the cost of carry and u_{ct} is a stationary disturbance term with mean zero and finite positive variance; and (3) the first difference of log-spot price Δs_t is a $I(0)$ process with mean zero and finite positive variance. With assumption (1) and (2), equation (1) can be converted to:

$$f_t - s_t = \bar{r} + \bar{c} + u_{rt} + u_{ct}. \tag{2}$$

It is generally accepted that under the standard no-arbitrage conditions, there are unit roots in both spot and futures prices (Dolatabadi et al (2016)) So, in equation (2), the two $I(1)$ processes, s_t and f_t are cointegrated to $I(0)$ (on the right-hand side), with cointegration vector (1, -1).

In empirical data of spot and futures prices, it is likely that there are costs of arbitrage, which makes the arbitrage riskier (Garbade and Silber (1983)). Figuerola-Ferretti and Gonzalo's (2010) model attributes the cost of arbitrage mainly to convenience yield, which is the benefit to the owner of the physical commodity (in this paper, Bitcoin) instead of its future contract. The convenience yield is denoted as $y_t = \gamma_1 s_t - \gamma_2 f_t$, where $\gamma_i \in (0,1)$ and $i = 1,2$. So, we have $f_t + y_t = s_t + r_t + c_t$. The equilibrium model of spot and futures prices with finite elasticity of arbitrage then becomes:

$$s_t = \beta_2 f_t + \beta_3 + u_{rt} + u_{ct}, \quad (3)$$

where the coefficient $\beta_2 = \frac{\gamma_2 - 1}{\gamma_1 - 1}$ and the constant $\beta_3 = \frac{\bar{r} + \bar{c}}{\gamma_1 - 1}$. When $\beta_2 > 1$, The futures price is lower than the spot price and tends to converge to it, which indicates long-run backwardation. This happens when there is positive convenience yield or benefits of owning the asset or commodity instead of its futures contract. When $\beta_2 < 1$, the futures price is higher than the spot price and moves towards it, which indicates long-run contango. It is beneficial to hold the futures contract instead of the asset or commodity. This situation usually happens when there is storage cost to hold a commodity or financial cost of carry to hold an asset. When $\beta_2 = 1$, there is no long-run backwardation or contango.

By introducing fractional integration, we can finally get the fractionally cointegrated equilibrium model (Dolatabadi et al (2016)). First, the assumption of the integration order for the above disturbance terms u_{rt} and u_{ct} changes from $I(0)$ to $I(1-b)$, with $b > \frac{1}{2}$ so that the processes are stationary and s_t and f_t are fractionally cointegrated. Second, the mean of Δs_t becomes μ instead of zero, allowing a drift in the spot price when $\mu \neq 0$. Then, Denote the new defined disturbance terms as v_{rt} and v_{ct}. The fractionally cointegrated equilibrium is:

$$s_t = \beta_2 f_t + \beta_3 + v_{rt} + v_{ct}. \tag{4}$$

2.2 The FCVAR Model Specification

The FCVAR model is first proposed by Johansen (2008) based on his study of CVAR model (Johansen (1995)). Through introducing the fractional integration order d and the degree of fractional cointegration b, the FCVAR model is:

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t \tag{5}$$

(see more details at Dolatabadi et al (2016)). In the formula, Δ^d is the fractional difference operator and $L_b = 1 - \Delta^b$ is the fractional lag operator $(d \geq b > 0)$. X_t is a p-dimensional vector containing the Bitcoin spot prices and futures prices, such that $X_t = (s_t, f_t)'$. The vector α ($p \times r$ matrix) is the speed of adjustment to the long-run equilibrium, and $\alpha_\perp$ would be the price discovery. β ($p \times r$ matrix) consists of r columns of cointegration vectors, such that $\beta' X_t$ are the cointegrating combinations of the variables in the system. As for $\Gamma_i = (\Gamma_1, \dots, \Gamma_k)$, they are the parameters of short-run dynamics of the variables. The error term ε_t is the p-dimensional disturbance identically and independently distributed with the mean 0 and covariance Ω. I assume the fractional integration order $d = 1$ in this paper, as is commonly assumed for most financial assets (Dolatabadi et al (2016)). And I am going to estimate the fractional cointegration b, which is the reduction in integration order in linear combination of the spot and futures prices $\beta' X_t$ compared to the prices $X_t = (s_t, f_t)'$.

By introducing the restricted constant in Johansen and Nielsen (2012), and the unrestricted constant in Dolatabadi et al. (2014) to the FCVAR model, the general error correction form of the model becomes:

$$\Delta^{\mathrm{d}} X_t = \alpha \Delta^{d-b} L_b(\beta' X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^{\mathrm{d}} L_b^i X_t + \xi + \varepsilon_t, \tag{6}$$

where ρ is the restricted constant term that restrict the constant term in the model as $\alpha\rho'$. It stands for mean of the long-run equilibrium in this model as $E\beta' X_t + \rho' = 0$. Another parameter ξ is the unrestricted constant term that gives a rise to the linear deterministic trend in the model.

In the non-fractional CVAR model, $d = b = 1$. It is a special case of the FCVAR model where fractionality is not considered:

$$\Delta X_t = \alpha\beta' L X_t + \sum_{i=1}^{k} \Gamma_i \Delta L^i X_t + \varepsilon_t. \tag{7}$$

Denoting X_t as Y_t, I get the traditional CVAR model:

$$\begin{aligned} \Delta Y_t &= \alpha\beta' L Y_t + \sum_{i=1}^{k} \Gamma_i \Delta L^i Y_t + \varepsilon_t \\ &= \alpha\beta' Y_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \varepsilon_t. \end{aligned} \tag{8}$$

In the empirical analysis below, besides testing for price discovery, I will also compare the FCVAR model with CVAR model to see if FCVAR model better fits the data of Bitcoin spot and futures prices.

2.3 Estimation Methods

According to the model estimation and inference of Johansen and Nielsen (2012) and Nielsen and Popiel (2014), the FCVAR parameters are estimated using conditional maximum likelihood. The maximum likelihood estimator of $(b, \alpha, \Gamma_1 \dots \Gamma_k)$ follows an asymptotic normal distribution, while (ρ, β) is asymptotically normally distributed when $b \leq 1/2$ and asymptotically mixed normally distributed when $b \geq 1/2$. Thus, the likelihood ratio (LR) tests for all parameters $(b, \rho, \alpha, \beta, \Gamma_1 \dots \Gamma_k)$ have an asymptotically chi-squared distribution.

2.4 Hypothesis Testing

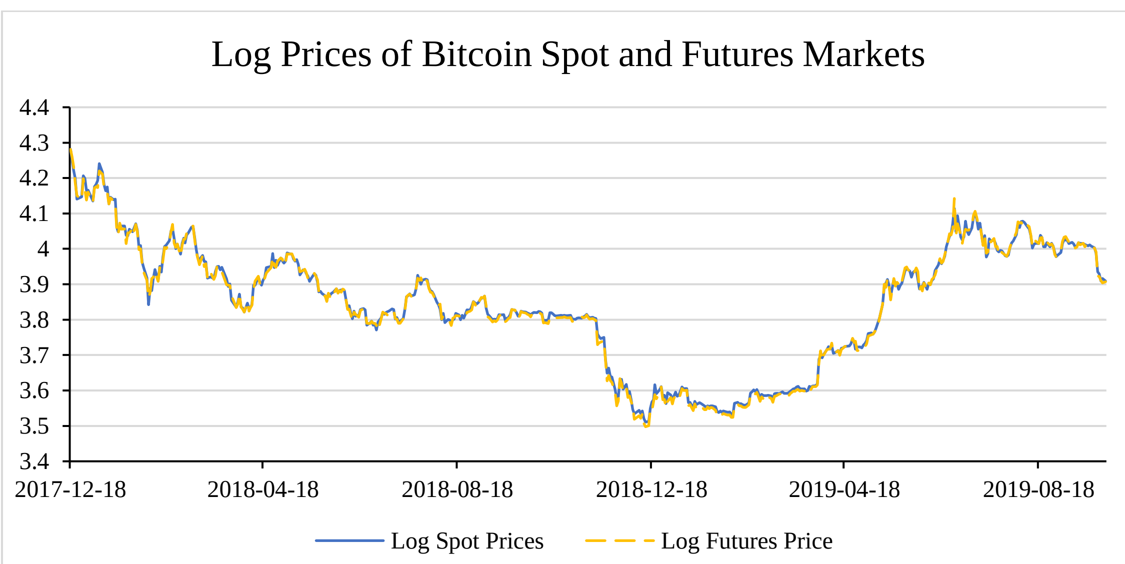
The general hypothesis for parameters estimating in FCVAR model is similar to the traditional CVAR model (Johansen (1995)). Of specific interest is the hypothesis on the price discovery parameter $\alpha_\perp$, which is orthogonal to the adjustment speed to the long-run equilibrium α. In the permanent-transitory (PT) decomposition of Gonzalo and Granger (1995), the permanent component of X_t is $\alpha'_\perp X_t$ where $\alpha'_\perp \alpha = 0$, which determines the fundamental value of the long-run equilibrium prices. According to Dolatabadi et al (2016), $\alpha_\perp$ is the price discovery parameter that stands for each market's contribution to the price discovery process. The null hypothesis is to determine which market (spot or futures) dominates in the price discovery process. I can assume price discovery "exclusively" (Dolatabadi et al (2016)) happens in the spot market. As $X_t = (s_t, f_t)'$, the null hypothesis would be $H_0^1\colon\ \alpha_\perp = (a, 0)'$. Or I can assume its "mirror hypothesis" (Dolatabadi et al (2016)) $H_0^1\colon\ \alpha = (0, \psi)'$. Oppositely, if I assume price discovery happens in the futures market, the null hypothesis would be $H_0^2\colon\ \alpha_\perp = (0, a)'$ or $H_0^2\colon\ \alpha = (\psi, 0)'$, where $\alpha = A\psi$, A is a $p \times m$ matrix of the restrictions, ψ is a $m \times r$ matrix $(m \geq r)$ of free parameters, and the degrees of freedom are $df = (p - m)r$. In the assumption that price discovery happens in the spot market, $A = (0, 1)'$. There is one restriction on α, so $df = 1$. The number of variables $p = 2$. The cointegration rank r equals 1, which is determined in section 3.3 below. So, $(p, m, r) = (2, 1, 1)$, and ψ is a free varying scalar.

3. Empirical Analysis

3.1 Data

The data used is daily and starts on 18 December 2017, which corresponds to the start of the issuance of Bitcoin futures. The end date for the data is 29 September 2019. The data series is for Bitcoin spot and futures prices. The spot price data is sourced from CoinMarketCap.com. (n.d.), which takes the volume-weighted average among Bitcoin prices among all market data found on the site. The futures price data is sourced from Intesting.com. (n.d.) (one dataset for each market), which uses the Chicago Mercantile Exchange (CME) market makers reported data. Bitcoin futures contracts expire three months in the future. The reason to use the two data sets is due to availability. The trading hours of CME Bitcoin futures are from Sunday to Friday from 6:00 p.m. to 5:00 p.m. (5:00 p.m. to 4:00 p.m. CT). To align the daily spot prices with futures prices, I have taken away the spot price data on Saturdays. There are 551 observations after the process.

Figure 2



Notes: Series are daily data from Dec 18, 2017 to Sep 29, 2019 except Saturdays. They are retrieved from https://coinmarketcap.com and https://www.investing.com. Data vintage is September 2019.

I can observe from Figure 2 that the Bitcoin spot prices and futures prices are evidently cointegrated as they are moving almost synchronously (Baur and Dimpfl (2018)).

3.2 Stationarity of Data

I conduct augmented Dickey-Fuller (ADF) tests and do not reject the hypothesis that there is a unit root in both price time series. In Table 1 and Table 2, the “t” statistics are greater than critical values at all conventional significance levels and the probabilities show the unit root hypothesis is not rejected for both logarithm futures and spot price time series. The non-stationarity of the time series allows further tests for the cointegration.

Table 1

Null Hypothesis: LOG FUTURES PRICE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.519545	0.1113
Test critical values:	1% level	-3.442011	
	5% level	-2.866576	
	10% level	-2.569512	

*MacKinnon (1996) one-sided p-values.

Table 2

Null Hypothesis: LOG_SPOT_PRICES has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.231145	0.1955
Test critical values:	1% level	-3.442054	
	5% level	-2.866595	
	10% level	-2.569522	

*MacKinnon (1996) one-sided p-values.

3.3 Test for Cointegration Rank

First of all, I need to decide the lag length in the FCVAR model. Use N=12 initial values (as the futures contracts are traded from Sunday to Friday each week, I pick 2-week data) for conditioning according to the initial values for conditional maximum likelihood estimation in Johansen and Nielsen (2014).

In table 3, the first panel reports a summary of model specification. In the second panel, the first column is the list of lags (k). The number of variables in the model (r) is 2 and the fractional integration order (d) equals 1 as mentioned in section 2.2. The reported values of the cointegration order (b) are estimated for each lag at the rank (r) and the fractional integration order (d). In the fifth column, there are the Log-likelihood (LogL) statistics for each lag. Under the null hypothesis $\Gamma_k = 0$, the likelihood ratio (LR) statistic and p-values (pv) are reported in the following 2 columns. After the AIC and BIC information criteria, the remaining columns report p-values for white noise tests on the residuals. The first p-value is for the multivariate Q-tests (pmvQ). Then pQ1 and pLM1 stand for p-values of univariate Q-tests and LM tests for the residuals in the first equation. pQ2 and pLM2 stand for the same tests for the residuals in the second equation.

The lowest value of AIC information criteria indicates the lag length of 2 may be a proper choice. Then look at the LR statistic and its p-value, which shows Γ_2 is significant by rejecting the null hypothesis of $\Gamma_2 = 0$ at 10% significance level. The corresponding fractional cointegration order b is greater than 1/2 and the large p-values of Ljung-Box Q tests show that I do not reject the null of white noise (no serial correlation). All these criteria indicate that Lag 2 (in bold) is the proper choice for the model.

Table 3

Lag Selection Results

Dimension of system:					2	Number of observations in sample:						551	
Order for WN tests:					12	Number of observations for estimation:						539	
Restricted constant:					Yes	Initial values:						12	
Unrestricted constant:					Yes	Level parameter:						No	
k	r	d	b	LogL	LR	pv	AIC	BIC	pmvQ	pQ1	pLM1	pQ2	pLM2
2	2	1	0.697	2404.66	8.12	**0.087**	**-4775.32**	-4702.40	0.72	0.82	0.95	0.34	0.54
1	2	1	0.756	2400.60	12.65	0.013	-4775.20	-4719.43	0.48	0.74	0.82	0.25	0.37
0	2	1	0.723	2394.27	0.00	0.000	-4770.55	-4731.94	0.16	0.60	0.74	0.29	0.49

After determining the lag length of the model, I can then decide the cointegration rank (number of cointegration relationships in the model). Table 4 and Table 5 reveals the process of cointegrating rank testing. The first panel in Table 4 reports the model specification. In the second panel of it, the first column is the list of ranks. The fractional integration order $d = 1$ is the same as above. The cointegration order b is estimated for each rank. Then the log-likelihood and likelihood ratio (LR) tests for each rank are presented in the following columns. The LR p-values in the last column are missing so I provide table 5 with likelihood ratio critical values.

The table needs to be read from the lowest rank to the highest rank. Clearly, I can reject the null of rank 0 against the rank 2 as its LR statistic is greater than the critical value at all conventional significance levels. Then test the null of rank 1 against rank 2. Since the LR statistic of rank 1 is smaller than the critical value at all conventional significance levels, I fail to reject the null. As expected, the cointegration rank equals 1, which means there is one cointegrated long-run equilibrium relationship between the spot and futures prices.

Table 4

Cointegrating Rank Testing

Dimension of system:	2	Number of observations in sample:	551
Number of lags:	2	Number of observations for estimation:	539
Restricted constant:	Yes	Initial values:	12
Unrestricted constant:	Yes	Level parameter:	No

rank	d	b	Log-Likelihood	LR statistic	p-value
0	1	0.697	2390.992	27.339	----
1	1	0.734	2403.127	3.069	----
2	1	0.010	2404.662	----	----

Table 5

Likelihood Ratio Critical Values

Unrest.cons	d	rank	b	C.V 0.01	C.V 0.05	C.V 0.1
1	1	2	0.01	0.000	0.000	0.000
1	1	1	0.73	9.443	6.113	4.669
1	1	0	0.70	19.901	15.621	13.499

3.4 Estimation of the FCVAR model

Having determined the lag length and cointegration rank in the model, I can then turn to the estimation of the FCVAR parameters. As is shown in the upper panel of Table 6, in the unrestricted model, the speed of adjustment coefficient $\hat{\alpha} = [-0.567, 0.854]'$, which shows the spot price $\hat{\alpha}_1$ is adjusting towards the equilibrium with its negative sign. The standard errors shown in the parentheses below are relatively large. So, the magnitude of the efficient estimates may not be very accurate. The unrestricted price discovery estimation $\hat{\alpha}_{\perp} = [0.601, 0.399]'$ is normalized so that each entry represents the contribution proportion of the spot and futures price. It indicates that the Bitcoin spot price dominates in the price discovery process with a 60% contribution. The cointegration order $\hat{b}$ is greater than 1/2, which shows the spot and futures prices are moving together as a stationary process.

Then I am going to estimate the restricted model using the likelihood ratio tests in the lower panel of Table 6. In the hypothesis testing, I have rejected the null hypothesis of $H_0: \beta = [1, -1]'$ in the first column for the p-value 0.003. It indicates that there is long-run backwardation or contango in the long-run relationship between Bitcoin spot and futures prices. Next, in the second column, I test the null hypothesis $H_0: b = d = 1$ which is to see if the CVAR model fits the data better than the FCVAR model. I reject the null of CVAR because of the zero p-value. Even though the FCVAR estimation of the spot market's contribution to the price discovery (60%) is lower than that in the CVAR model (91%), the hypothesis testing result indicates the necessity to include the fractional parameters in the model. The FCVAR model is more appropriate in estimating the price discovery of Bitcoin spot and futures markets. In the third column, I test the null hypothesis that price discovery happens exclusively in the spot market (or the futures market has no contribution to price discovery). The p-value shows I do not reject the null. In the last column, reversely, I test the null hypothesis that price discovery happens exclusively in the futures market (or the spot market has no contribution to price discovery). The p-value shows I do not reject the null at the 10% significance level but reject at the 1% and 5% significance level. So, I conclude that the spot market dominates in price discovery of Bitcoin.

Table 6

Unrestricted FCVAR Model			
Unrestricted Model			
$\hat{\alpha} = [-0.567, 0.854]'$ (0.487) (0.423)	$\hat{\alpha}_\perp = [0.601, 0.399]'$	$\hat{b} = 0.735$ (0.100)	
Hypothesis Testing			
$H_0: \beta = [1, -1]'$	$H_0: b = d = 1$	$\boldsymbol{H_0^1: \alpha_\perp = [a, 0]'}$	$H_0^2: \alpha_\perp = [0, a]'$
$\hat{\alpha}_\perp = [0.509, 0.492]'$	$\hat{\alpha}_\perp = [0.912, 0.088]'$		
$df = 1$	$df = 1$	$df = 1$	$df = 1$
LR $= 9.046$	LR $= 15.648$	LR $= 1.940$	LR $= 3.420$
P-value $= 0.003$	P-value $= 0.000$	**P-value $= \mathbf{0.164}$**	P-value $= 0.064$

In Table 7, I estimate the restricted FCVAR model when price discovery happens exclusively in the spot market, for which $\hat{\alpha}_1 = 0$. The fractional cointegration order $\hat{b}$ is greater than 1/2, indicating the cointegrated process of Bitcoin spot and futures prices is stationary. $\hat{\beta}_2$ is smaller than 1, indicating there is long-run contango. The equilibrium relationship between spot and futures prices is shown at the bottom of the table. The futures price is higher than the spot price and converges to it. The convenience yield is negative, which implies there is a low demand for Bitcoin at the spot market (Figuerola-Ferretti and Gonzalo (2010)) or traders expect the Bitcoin price in the future to be higher than the present. Figure 3 depicts the Bitcoin Spot prices, the 3-month futures price, and futures-3-month contango defined as $F_t - S_t$(Figuerola-Ferretti and Gonzalo (2010)). The contango shows more volatile changes when the Bitcoin spot and futures prices are high, which can be the result of a large trading volume. Recall Figure 1, there is a drastic increase in Bitcoin trading volume since 2016. It reaches the pick in December 2017, declines to the bottom in December 2018, and then recovers to a new relatively lower peak in July 2019. The increase in the demand for bitcoin leads to an increase in the price in Figure 3. According to Figuerola-Ferretti and Gonzalo (2010), contango can be constraint by adding up inventory (here it can be the increased demand of Bitcoin). So, Figure 3 shows a negative volatile contango (backwardation) in December 2017 and it becomes more positive after the price declines (which implies demand also declines). When demand and price recover in 2019, the constraint on contango also recurs.

Table 7

Restricted FCVAR model

Restricted Model

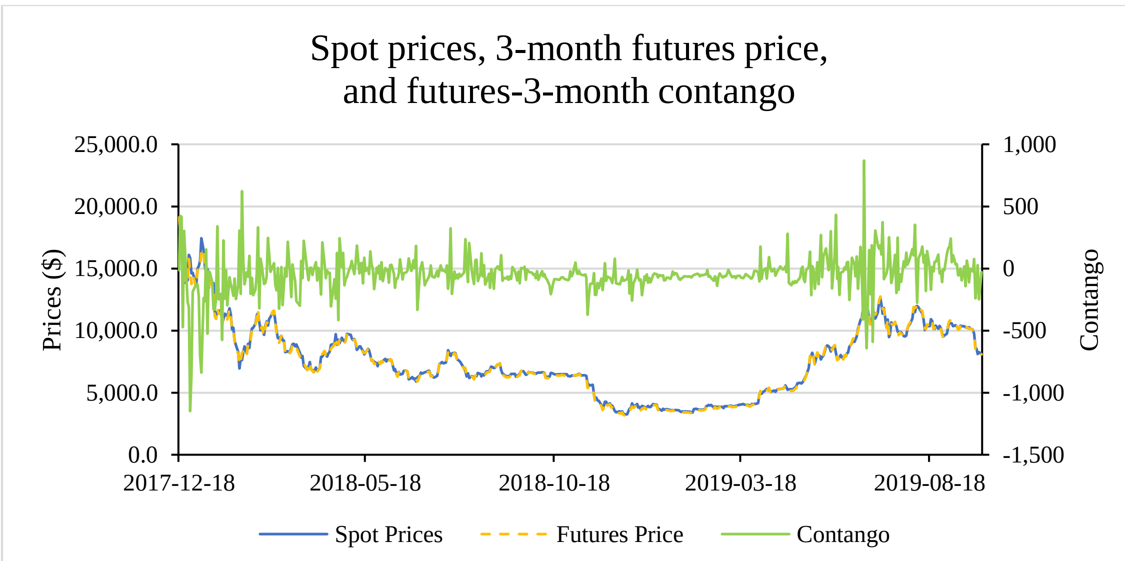
$\hat{\alpha} = [0.000, 1.308]'$ (0.000) (0.347) $\quad \hat{b} = 0.762$ (0.092) $\quad \hat{\beta} = [1.000, -0.987]'$ $\quad \hat{\rho} = -0.108$

$$\Delta \begin{bmatrix} s_t \\ f_t \end{bmatrix} = \hat{\alpha}\Delta^{1-\hat{b}} L_{\hat{b}} \left(\hat{\beta}' \begin{bmatrix} s_t \\ f_t \end{bmatrix} + \hat{\rho}'\right) + \sum_{i=1}^{2} \Gamma_i L_{\hat{b}}^{i} \Delta \begin{bmatrix} s_t \\ f_t \end{bmatrix} + \begin{bmatrix} 0.001 \\ -0.005 \end{bmatrix} + \hat{\varepsilon}_t$$

Equilibrium relationship

$$s_t = -0.108 - 0.987 f_t + v_t$$

Figure 3



Notes: Series are daily data retrieved from https://coinmarketcap.com and https://www.investing.com.

Data vintage is September 2019.

4. Conclusions

In this paper, I studied the price discovery in the Bitcoin spot and futures markets using the fractionally cointegrated vector autoregressive (FCVAR) model. Through the PT decomposition, I can decompose the observed price into a permanent component and transitory component. The price discovery process is just the orthogonal component of the adjustment speed to the permanent price, that is the long-run equilibrium price of the Bitcoin. According to the spot-futures parity condition, the spot price and futures price has a cointegrated long-run equilibrium relationship based on the cost of carry.

The reason why I choose to adopt the FCVAR model is its superiority over the traditional cointegrated vector autoregressive (CVAR) model. FCVAR model introduces fractional parameters in the model formulation, while the CVAR model only allows integers. It implies that the FCVAR model can provide more precise estimations and more information on the empirical data. This paper has also proved that the fractional parameter is significant in the empirical analysis, which becomes the evidence of the appropriateness of adopting this method.

In the research of the price discovery, I conclude that the spot market dominates in the price discovery for Bitcoin, which may be the result of much larger transaction volume in the Bitcoin spot market than in the futures market. It's reasonable as there are more participants and information in the market, the price discovery will be faster. My result is opposite from the result of Hasbrouck (1995) information share model and Gonzalo and Granger's (1995) common factor component but consistent with the CVAR analysis in the literature. As the price discovery leadership is closely related to the market size, liquidity and, trading volume, the dominance of the spot market may switch to the futures market when those conditions change.

References

Baur, D. G. & Dimpfl, T. (2018). Price Discovery in Bitcoin Spot or Futures? Available at SSRN: https://ssrn.com/abstract=3171464 or http://dx.doi.org/10.2139/ssrn.3171464

Brandvold, M., Molnár, P., Vagstad, K., & Valstad, C. A. (2015). Price discovery on Bitcoin exchanges. *Journal of International Financial Markets, Institutions & Money, 36*, 18–35.

Brauneis, A. & Mestel, R. (2018). Price discovery of cryptocurrencies: Bitcoin and beyond. *Economics Letters*. 165, 58-61.

Chen, Y. & Gau, Y. (2010). News announcements and price discovery in foreign exchange spot and futures markets. *Journal of Banking & Finance*, 34, 1628–1636.

Chiu, J. & Koeppl, T. V. (2019). The Economics of Cryptocurrencies—Bitcoin and Beyond. Retrieved January 13, 2020, from https://www.bankofcanada.ca/2019/09/staff-working-paper-2019-40/. Sep.

CoinMarketCap.com. (n.d.). Retrieved September 25, 2019, from https://coinmarketcap.com/currencies/bitcoin/historicaldata/?start=20180922&end=20190922

De Jong, F., Mahieu, R., Schotman, P., van Leeuwen, I. (2001). Price discovery on foreign exchange markets with differentially informed traders.

Dolatabadi, S., Narayan, P. K., Nielsen, M. Ø., & Xu, K. (2015). *Economic significance of commodity return forecasts from the fractionally cointegrated VAR model.* the Conference on Recent Developments in Financial Econometrics and Applications, Centre for Economics and Financial Econometrics Research, Deakin University, Australia. July.

Dolatabadi, S., Nielsen, M. Ø., & Xu, K. (2016). *A fractionally cointegrated VAR analysis of price discovery in commodity futures markets.* The Conference on Performance of Financial Markets and Credit Derivatives, Centre for Economics and Financial Econometrics, Deakin University, Australia. Feb.

Dolatabadi, S., Nielsen, M. Ø., & Xu, K. (2016). A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets. *Journal of Empirical Finance*. 38, 623–639.

Figuerola-Ferretti, I. & Gonzalo, J. (2010). Modelling and measuring price discovery in commodity markets. *Journal of Econometrics*. 158, 95–107.

Foster, A. J. (1996). Price discovery in oil markets: a time varying analysis of the 1990-91 Gulf conflict. *Energy Economics*, 18, 231-246.

Garbade, K. D. & Silber, W. L. (1983). Price Movements and Price Discovery in Futures and Cash Markets. *The Review of Economics and Statistics*. 65, 289-297. The MIT Press. May.

Gonzalo, J. & Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*. 13, 27-35. Jan.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*. 16, 121-130.

Hasbrouck, J. (1995). One security, many markets: determining the contributions to price discovery. *The Journal of Finance*. 50, 1175–1199. Sep.

Intesting.com. (n.d.). Retrieved September 25, 2019, from https://www.investing.com/crypto/bitcoin/bitcoin-futures-historical-data

Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. New York: Oxford University Press.

Johansen, S. (2008). A Representation Theory for a Class of Vector Autoregressive Models for Fractional Processes. *Econometric Theory*, 24, 651-676. Jun.

Johansen, S. & Nielsen, M. Ø. (2012). Likelihood Inference for A Fractionally Cointegrated Vector Autoregressive Model. *Econometrica*, 80, 2667–2732. Nov.

Johansen, S. & Nielsen, M. Ø. (2014). The role of initial values in nonstationary fractional time series models. QED Working Paper 1300, Queen's University.

Kapar, B. & Olmo, J. (2019). An analysis of price discovery between Bitcoin futures and spot markets. *Economics Letters*, 174, 62-64.

Livni, E. (2019). Facebook's Libra is spurring central banks' interest in issuing cryptocurrency. Retrieved October 31, 2019, from https://qz.com/1655896/facebooks-libra-forces-central-banks-to-tackle-cryptocurrencies/. June.

Nielsen, M. Ø. & M. K. Popiel (2016), "A Matlab program and user's guide for the fractionally cointegrated VAR model," QED working paper 1330, Queen's University.

Kavussanos, M. G. & Nomikos, N. K. (2003). Price Discovery, Causality and Forecasting in the Freight Futures Market. *Review of Derivatives Research*, 6, 203–230.