The Determinants of Price Discovery on Bitcoin Markets*

Oliver Entrop†

University of Passau

Bart Frijns‡

Auckland University of Technology

Marco Seruset§

University of Passau

Working Paper

July 2019

* Parts of this research was done while Oliver Entrop was visiting the Auckland University of Technology. He thanks Aaron Gilbert and the academic and administrative staff for their hospitality and support.

† Oliver Entrop, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2460, email: oliver.entrop@uni-passau.de

‡ Bart Frijns, Auckland University of Technology, Department of Finance, Private Bag 92006, 1142 Auckland, New Zealand, phone: +64 9 921 9999x5706, email: bfrijns@aut.ac.nz

§ Marco Seruset, University of Passau, Chair of Finance and Banking, Innstraße 27, 94032 Passau, Germany, phone: +49 851 509 2463, email: marco.seruset@uni-passau.de

The Determinants of Price Discovery on Bitcoin Markets

Abstract

This paper is the first to investigate whether market quality and uncertainty affect bitcoin price discovery in spot and futures markets. Using high-frequency data over the period December 2017 – March 2019, we find significant time variation in the contribution to price discovery of the two markets. Considering potential endogeneity issues between price discovery and measures of market quality, we document that increases in price discovery are mainly driven by relative trading costs and relative trading volume, and by uncertainty to a lesser extent. Additionally, we show that medium-sized trades contain most information in terms of price discovery.

Keywords: Bitcoin; Price Discovery; Futures.

JEL classification: G12; G13; G14

1 Introduction

Cryptocurrencies[1] and especially bitcoin[2] have received increasing attention in the academic finance literature in recent years. Much of this research focuses on issues such as long- and short-term determinants of the exchange value of bitcoin (e.g., Kristoufek, 2015; Li and Wang, 2017; Mai et al., 2018), the market efficiency of bitcoin (e.g., Urquhart, 2016; Köchling et al., 2018), the diversification effects and connectedness of bitcoin with other financial assets (e.g., Brière et al., 2015; Dyhrberg, 2016; Bouri et al., 2017; Corbet et al., 2018), illegal activities (e.g., Foley et al., 2019), or the price discovery process among bitcoin trading venues (e.g., Brandvold et al., 2015; Pagnottoni and Dimpfl, 2019).

In December 2017, the CME and CBOE introduced bitcoin futures, enabling investors to trade and hedge bitcoin on regulated markets. The introduction of this new market raises two important questions related to price discovery. First, which market, i.e., spot or futures, lead the bitcoin price discovery process?[3] Second, what are the determinants of price discovery? The first question has been the focus of three recent studies. Corbet et al. (2018), and Baur and Dimpfl (2019) explore price discovery leadership using high-frequency transaction data and find that the spot market incorporates information into prices first and thus dominates in terms of price discovery. In contrast, using daily data, Kapar and Olmo (2019) find that the futures market is the price discovery

[1] According to coinmarketcap.com, over 2,000 cryptocurrencies exist with a total market capitalization surpassing 172 billion US Dollar as of 11 April, 2019.

[2] A detailed description of the Bitcoin technology is provided in Nakamoto (2008), Kroll et al. (2013) and Boehme et al. (2015).

[3] This standard microstructure analysis between spot and futures markets has already been subject for various asset classes, such as stocks (e.g., Hasbrouck, 1995; Booth et al., 1999), exchange rates (e.g., Chen and Gau, 2010) and commodities (e.g., Dimpfl et al., 2017).

leader. To the best of our knowledge, the second question on determinants of price discovery has not been addressed yet.

Our study extends the existing literature in two important directions. First, while the studies mentioned above examine price discovery using the full contract term of each separate futures contracts, we consider the liquidity of each contract on each day. Specifically, we determine the daily contribution to price discovery based on the most actively traded futures contract, which allows us to capture the potential dynamics in the relation between spot and futures markets on a day-to-day basis. Using a sample of high-frequency midquotes over the period December 2017 to March 2019, this first-stage analysis demonstrates that price discovery in bitcoin markets is subject to time variation. Using the Gonzalo and Granger (1995) Component Share and Hasbrouck (1995) Information Share, we find that, on average, the futures market leads the price formation process in nine (contract) months, while the spot market is the leader in the remaining (six) months. In our robustness section, we further observe that the price discovery measures get closer to 0.5 when increasing time intervals. One of the critical points we raise in this stage is that the spot market does not lead the price discovery process exclusively.

Second, we analyze the effect of market quality, uncertainty, and other controls on daily price discovery. Frijns et al. (2015) argue that the relation between price discovery and measures of market quality, such as trading costs and trading activity, is potentially endogenous, where an enhancement in price discovery may attract investors to a market, while an increase in liquidity, trading activity, and lower trading costs may improve price discovery. We, therefore, implement 2SLS time-series regressions to control for potential endogeneity. Our results show that trading costs, captured by the relative bid-ask spread, are negatively associated with price discovery, while relative trading volume is positively related to price discovery. Thus, an increase in relative spread (relative trading volume) in one market relative to the other market, decreases (increases) the contribution to price discovery of that market. Quoting activity does not affect price discovery. Furthermore, measures of uncertainty such as volatility of the spot market and VIX partially reveal a significant shift of price discovery to the futures market. Beyond that, we find in additional analyses that the relative number of medium-sized trades contains most information for the price discovery process.

Baur and Dimpfl (2019) point out that the analysis of bitcoin price discovery may be somewhat different from other asset classes. Given the absence of a bitcoin pricing model, the ambiguity to which asset class the bitcoin even belongs to, as well as the different design of spot (unregulated) and futures markets (regulated), one ex-ante cannot expect that the results of other asset classes also hold for the bitcoin market. Though the time variation in price discovery we observe in our first stage is in line with the findings in the DAX ETF and DAX futures market (see Schlusche, 2009), and in the VIX short-term futures ETN and inverse VIX short-term ETN (see Fernandez-Perez et al., 2018). In contrast, studies on price discovery between spot and futures markets, often find the futures market to lead (see, e.g., Chen and Gau, 2010, for foreign exchange spot and futures markets; Theissen, 2012, for the DAX spot and DAX futures; Dimpfl et al., 2017, for spot and futures of eight agricultural commodities). In our second-stage analysis, we observe a significant effect of trading volume and trading costs on price discovery. This is consistent with other studies that have also focused on the relation between market quality and price discovery on spot and derivatives markets (see, e.g., Chakravarty et al., 2004, for stocks and stock option markets; Fernandez-Perez et al., 2018, for VIX short-term futures ETN and inverse VIX short-term ETN). Our results concerning uncertainty suggest that the relative contribution of the futures market to price discovery is higher when volatility on the bitcoin spot market and stock markets is higher.

For spot market volatility, our findings are in contrast to the stocks and stock options markets (see Chakravarty et al., 2004), but in line with the foreign exchange spot and futures markets (see Chen and Gau, 2010). The mechanism of the VIX relating to bitcoin price discovery is difficult to assess and has not been addressed in such a setting. Overall, our findings imply that the price discovery on bitcoin markets are not too different from other asset classes.

The remainder of this paper is organized as follows. Section 2 describes the data and presents summary statistics. In Section 3, we present the model used to evaluate price discovery, present our empirical results, and discuss several robustness tests. Section 4 focuses on the determinants of price discovery and reports results of our second-stage analysis. We conclude in Section 5.

2 Data

This study concentrates on the dynamic relation between bitcoin spot and futures prices from December 17, 2017 to March 31, 2019. We consider intraday trade and quote data for bitcoin futures traded on Chicago Mercantile Exchange (CME) as well as the corresponding spot of the Bitstamp (BTSP) exchange. We obtain these data from the Thomson Reuters Tick History (TRTH) database.[4]

The transaction data include the timestamp to the nearest millisecond, the traded price, and associated volume. The quote data consist of the bid and ask quotes, and the exact timestamp a new quote is issued. From this, we calculate the midpoint (average of bid and ask quotes) for spot and futures.

[4] Note that we do not consider futures contracts traded on Chicago Board Option Exchange (CBOE). First, CBOE has announced that bitcoin futures will no longer be listed as of March 2019. Second, notional trading volume on CME is superior to CBOE from March 2018 onwards. Therefore, we assume the CME to be the relevant futures market.

CME bitcoin futures (RIC: BTC) are US dollar-denominated cash-settled contracts, based on the CME CF Bitcoin Reference Rate (BRR), having a contract size of five bitcoins. The BRR aggregates the weighted median USD price for four major exchanges (Bitstamp, Coinbase, itBit, and Kraken) once a day. Trading in expiring futures contracts terminates at 4 pm London Time on the expiration day. The trading hours for CME futures contracts are between 5 pm and 4 pm Chicago Time (CT) from Sunday to Friday with a 60-minute break each day beginning at 4 pm CT.[5]

We follow Baur and Dimpfl (2019) and select the Bitstamp spot as the spot price. (We do not use the daily available Bitcoin Reference Rate (BRR) nor its continuous version (Bitcoin Realtime Index – BRTI) because investors cannot trade these indices). Bitstamp is one of the largest cryptocurrency spot trading platforms, where bitcoin can be traded against USD (RIC: BTC=BTSP).[6]

The analysis of the daily behavior of price discovery requires a continuous futures time series. We follow Fricke and Menkhoff (2011) and Hauptfleisch et al. (2016) and use the most actively traded futures contract on each day in our sample. An alternative procedure in empirical studies is to use the nearest-to-maturity futures contract (e.g., Booth et al., 1999; Cabrera et al., 2009). In our case, however, there are only minor differences when comparing the time series resulting from both methods. In particular, the most actively traded futures contract equals the nearest-to-maturity contract until one business day before maturity. At that point, volume shifts to the second-nearby contract, implying that the closest-to-maturity contract is no longer the most actively traded.

[5] See https://www.cmegroup.com/trading/equity-index/us-index/bitcoin_contract_specifications.html for more details.

[6] See https://www.bitstamp.net/ for more information.

Another important issue of data preparation relates to the trading hours of the futures contracts. Similar to Grammig et al. (2005), we consider overlapping trading hours between spot and futures only. We further follow the procedure of Hauptfleisch et al. (2016) and delete all entries before 0 am and after 8 pm GMT. This avoids the need to deal with market closures on CME and time zone transformations, thus simplifying our two-stage analysis. Finally, we remove all observations on holidays according to CME holiday calendar.

[Table 1 about here]

Column 2 of Table 1 shows the time interval in which the respective futures contract (RIC) is the most actively traded. Column 4 presents the total daily volume of the most-traded futures (MTF) in the respective time period. Interestingly, volume increases nearly monotonically until August 2018, while we observe a more volatile behavior of volume after August 2018 until the end of the sample. Column 5 emphasizes the importance of using the most actively traded futures contracts for analyzing the dynamic price discovery process. For example, BTCQ8 exhibits an average proportion of 96.65%, indicating that there is almost no trading in other contracts at that time. This strong shift in liquidity between futures contracts may favor previous empirical results of spot-driven price discovery (e.g., Corbet et al., 2018; Baur and Dimpfl, 2019) when futures contracts are considered over their whole life span.

Finally, the analysis of price discovery between spot and futures can be conducted on either quotes or transaction prices. Several studies have already discussed the advantages of using midquotes over transactions data (see, e.g., Shyy et al., 1996; Eun and Sabherwal, 2003; Grammig et al., 2005; Theissen, 2012). The use of quote midpoints implies three main advantages. First, quotes can be updated in the absence of transactions. Second, midquotes mitigate the problem of infrequent trading, which is normally observed in transaction prices. Third, midquotes are not affected by the bid-ask bounce. Hence, we base our analysis on midquotes.

We estimate the contribution to price discovery of the spot and futures separately for each day in our sample period to capture the dynamic behavior of the price formation process. Since midquotes of bitcoin spot and futures are not uniformly spaced in time, we construct synchronized time intervals to align the spot and futures data. Within each time interval, we keep the last observed midquote. If no midquote is observed, we fill missing intervals with the most recent non-missing value (see, e.g., Chan, 1992; Chen and Gau, 2010).[7] The choice of sampling interval is an important issue when studying price discovery. Brandvold et al. (2015) and Jin et al. (2018) point out that it is important to keep time intervals short enough to ensure information is not lost between sampling intervals, but also long enough to avoid noise due to stale prices. Following Jin et al. (2018), we consider various sampling frequencies in our analysis. In particular, we compute the non-synchronous quoting probability, as well as the frequency of zero-returns as zero-returns are an important indicator of liquidity differences between spot and futures markets (Theissen, 2012). It should be noted, however, that different trading activity and different liquidity does not necessarily have to be an indication of the leading market (see, e.g., Theissen, 2012; Jin et al., 2018).

Table 2 reports the trading frequency and the proportion of zero-returns. We observe a lower proportion of missing quotes on the spot market. On average, the non-synchronous quoting for one-minute intervals is 0.35% and 4.40% for the spot and futures market, respectively. Non-synchronous quoting decreases as we increase the time interval. When we consider the proportion

[7] For an alternative procedure of constructing a matched sample of midquotes see Harris et al. (1995).

of zero returns, however, figures substantially increase. Zero returns for spot and futures prices occur in 15.32% and 43.37% of the one-minute return intervals, respectively. Thus, midquotes change more frequently in the spot market than in the futures market. We proceed with our price discovery analysis using one-minute intervals, but also consider five-, ten- and fifteen-minute intervals for robustness purposes in our first stage.

[Table 2 about here]

Table 3 presents summary statistics for one-minute intervals based on midquotes. The average quote midpoint is 7,035 for spot and 7,031 for futures. Bitcoin spot and futures midquotes show a declining trend, which results in a negative return of almost 80% from the start to the end of our sample period.

The non-synchronicity between spot and futures is remarkably low for all contracts in our sample, which again supports our decision to analyze price discovery on a one-minute frequency. However, figures increase when we consider the evolution of zero returns, where futures always exhibit a higher percentage of zero returns than the spot. In terms of percentage changes, however, the pattern is not uniform over the sample period. The percentage of zero returns increases fivefold between the January (BTCH8) and the June contract (BTCM8) for spot and futures. In the subsequent contract months, the percentage of zero returns increase for the spot market, while the futures market's zero returns decrease. After the September contract (BTCU8), the spot and futures market reveal nearly a doubling in the percentage of zero returns until March 2019 (BTCH9). The growth in the zero returns is more volatile than before.

[Table 3 about here]

3 Price Discovery

To study the dynamics of the price discovery process between bitcoin spot and futures prices, we apply the standard approach of estimating a vector error correction model (VECM) and deriving our price discovery measures directly from the outcome of the VECM. We use two of the most important price discovery measures for non-stationary price series developed by Gonzalo and Granger (1995), i.e., Component Share (CS), and Hasbrouck (1995), i.e., Information Share (IS). Subsequently, we present the results of the VECM as well as the price discovery measures.

3.1 Vector error-correction model and price discovery measures

We are interested in questions related to the intra-day relation between bitcoin spot and futures prices. Suppose Bitstamp spot has a log US dollar price s_t, and f_t denotes the log US dollar price of the CME futures. Let $\mathrm{y}_t = (s_t \quad f_t)'$ be the vector of these price series. Given the cost-of-carry relation between spot and futures prices, the respective log price series should be integrated of order one, $I(1)$, with cointegrating vector $\beta' = (1 \quad -1)$ (see Baur and Dimpfl, 2019). Therefore, price changes can be expressed as an error correction equation of the form

$$\Delta \mathrm{y}_t = \alpha(\beta' \mathrm{y}_{t-1} + \mu) + \sum_{i=1}^{p} \Gamma_i \Delta \mathrm{y}_{t-i} + \varepsilon_t, \tag{1}$$

where $\Delta \mathrm{y}_t$ is the (2 x 1) vector of changes in the log series of the spot and futures price at time t. α is a (2 x 1) vector for the bitcoin spot and futures prices measuring the speed of adjustment of short-term deviations from the long-term equilibrium. Our specification of β' implies that we expect $\alpha^{Spot} \leq 0$ and $\alpha^{Futures} \geq 0$. μ is a constant term[8] in the cointegrating equation, and Γ_i are

[8] Note that this constant term refers to the restricted constant specification as defined by Johansen (1995). According to Hansen and Juselius (1995) this is the minimum deterministic component recommended by Johansen

(2 x 2) matrices of autoregressive prices, representing the short-term transitory effects due to market imperfections. ε_t is a zero-mean vector of serially uncorrelated innovations with the following covariance matrix:

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{2}$$

where σ_1^2 (σ_2^2) is the variance of spot market innovations (futures market innovations) and ρ is the correlation between these innovations.

Appendices A and B outline the calculation of the Component Share (CS) and Information Share (IS) from the outcome of Equation (1). Values above (below) 0.5 suggest that the spot (futures) market leads the price formation process.

Frijns et al. (2015) point out that the IS may be biased when liquidity increases over time. In such a case, a rise in liquidity increases the contemporaneous correlation and widens the lower and upper bound. This bias causes the IS to move towards 0.5 for both markets.[9] Indeed, we observe that liquidity of the spot and futures market has changed over time (see Section 2). For this reason, we calculate the CS and IS for each day in our sample, but focus only on the CS in our second-stage analysis.

(1995). This allows the cointegrating equations to be stationary around a constant mean, which seems appropriate for daily estimation procedure. We conduct all analysis based on this specification pointing out, however, that our results are robust to the choice of the deterministic component, i.e. results of price discovery remain qualitatively and quantitatively the same for constant or restricted trend specifications. For a more detailed discussion, see Ahking (2002).

[9] For a numerical example, see Putniņš (2013). As the Information Share Leadership (see Yan and Zivot, 2010; Putniņš, 2013) is also affected by this problem, we do not consider this measure.

3.2 Empirical analysis

Analyzing the price discovery process of two time series requires data to be cointegrated. For this purpose, we determine the number of cointegrating equations by Johansen's (1995) trace statistic method.[10] We determine the lag length included in the model by the multivariate version of Schwartz's Bayesian Criterion (SBIC).[11] Our first step is to test whether there are at most zero cointegrating vectors for each day in our sample. The null hypothesis of $r = 0$ cointegrating vectors is rejected for around 91% of the days at the 1% level. In the next sequence, the null hypotheses of $r = 1$ cointegrating vectors cannot be rejected for about 79% of those days. We thus discard 21% of days from our data set. The mean cointegrating equation is $\beta' = (1 \quad -0.89307)$. However, we cannot reject the null hypothesis that the cointegrating relation is $\beta' = (1 \quad -1)$ at the 5% level.[12]

We confirm the presence of one cointegrating relation on almost 80% of the days in our sample. Our next aim is to investigate the price discovery dynamics between bitcoin spot and futures using two measures of price discovery, the Component Share (CS) and the Information Share (IS). Once again, it is worth noting that the results of price discovery refer to the spot market and that values above (below) 0.5 indicate that the spot (futures) market is the leading market.

Table 4 reports the CS and IS for each most-traded futures (see Table 1) in our sample, based on one-minute intervals (Panel A). We document that the futures market leads the spot market in

[10] We additionally perform unit root tests for both series for each day in our sample. Results of Augmented Dickey-Fuller tests for the log-levels of spot and futures reveal that roughly 82% of the days are non-stationary (at the 1% level), while first differences are always stationary.

[11] The average lag length for each day is $p = 3$.

[12] For detailed results of the VECM estimation see Table A1 in the Appendix. By definition of the VEC model stated in (1), β_{Spot} is 1 and, by theory, $\beta_{Futures}$ is –1. Due to outliers in beta estimations, we observe that the mean beta significantly deviate from the theoretical value in contract months M8 and F9. Additionally, t-values of beta estimates are significant at the 1% level in six out of fifteen contract months, indicating that the cointegrating vector does not hold. These indistinct results, however, are in line with the findings of Baur and Dimpfl (2019). The median value turns out to be the better indicator in this case, where we observe a reasonably tight range of median figures. Therefore, we assume that the theoretical cointegrating equation $\beta' = (1 \ -1)$ holds for all days in our sample.

nine contract months (price discovery measures < 0.5), while three months are significant at the 1% level (Column 2). The spot market is the leading market in the remaining months with two significant months (price discovery measures > 0.5). Over the full sample period (Panel B), we, on average, observe that price discovery measures are close to 0.5. Overall, the Information Share produces similar results with respect to the price discovery leader, however, with one more significant contract month at the 5% level (BTCQ8). In summary, the importance of spot and futures market in incorporating new information changes over time. The variability is also visualized by the 5-day moving average in Figure 1, which is calculated from the daily CS and IS. We, again, point out that the time variation in price discovery can also be observed in other asset classes, such as DAX ETF and DAX futures market (see Schlusche, 2009) as well as in the VIX short-term futures ETN and inverse VIX short-term ETN (see Fernandez-Perez et al., 2018).

[Table 4 about here]

[Figure 1 about here]

Considering the distributional properties of the price discovery measures over the contract months (Table 4, Panel A), CS compared to IS, is more volatile and reveals a wider difference between the 95th and the 5th percentiles. Moreover, on average, we observe a lower standard deviation for significant contract months, ranging from 15.3% to 22.7% for CS, and from 6.4% to 12.2% for IS, respectively.

For robustness purposes, we replicate our analysis for five-, ten- and fifteen-minute intervals. Table 5 documents the CS and IS for the different sampling intervals. In line with Jin et al. (2018), price discovery shares get closer to 0.5 when lower-frequency intervals are used, on average. Stated differently, the differences in price discovery shares between the spot and futures market are less when increasing time-intervals (see Tse et al., 2006, for similar results). This fact confirms that information transmission between the spot and futures market takes less than fifteen minutes.

[Table 5 about here]

4 Determinants of price discovery

4.1 Potential determinants and summary statistics

In our second-stage analysis, we examine different variables that may explain our previous price discovery findings. For this purpose, we consider three sets of variables. These data are either calculated from the data as described in Section 2 or collected from Thomson Reuters Eikon.

Market Quality

The first set of variables capture various aspects of market quality, such as trading activity or trading costs of the bitcoin spot and futures market. Following earlier studies (e.g., Frijns et al., 2015; Fernandez-Perez et al., 2018), we consider the relative number of quotes ($rel\_num\_Quotes_t$), which is the number of quotes on the spot market divided by the number of quotes on the futures market on day t. We also take into account the relative trading volume ($rel\_vol\_trades_t$), which is the volume of contracts traded on the Bitstamp spot market divided by the volume of traded contracts on the CME futures market on day t. The variable rel_BAS_t is defined as the daily average percentage bid-ask-spread on the Bitstamp spot market divided by the daily average percentage bid-ask-spread on the CME futures market.

We also consider the relative size of each trade in a subsequent analysis. In particular, we decompose the relative traded volume into small, medium, and large trades. Large trades ($rel\_num\_large\_trades_t$) are those of five futures contracts[13] or five bitcoins, respectively, or more; small trades ($rel\_num\_small\_trades_t$) are defined with a respective number of less or equal one, while medium-sized trades ($rel\_num\_medium\_trades_t$) are those with a respective number of more than one and less than five.[14]

Uncertainty

Our second set of variables contains several measures of uncertainty. We include the Bitstamp spot market volatility ($rel\_vola_{t,Spot}$), which is defined as the square root of the sum of the squared 1-min returns for each day in our sample, similarly done by Chakravarty et al. (2004) and Chen and Gau (2010). This variable serves as a proxy of the uncertainty on the bitcoin market. We also include the daily log-return of VIX (ret_VIX_t), which is often used as a proxy of fear on stock markets, or even as a general fear measure for capital markets. In addition, we consider the economic policy index lagged by two periods (EPU_{t-2}; see Wang et al., 2014), which was developed by Baker et al. (2013) for the US. It serves as a proxy of real economic policy uncertainty.

Controls

[13] This boundary refers to the block trading limit of CME, where trades are negotiated manually between the exchange and investors. See https://www.cmegroup.com/education/bitcoin/cme-bitcoin-futures-frequently-asked-questions.html for more details.

[14] Note that the definition of different trading sizes is not homogenous in literature. Some researchers define the trading sizes according to the contract volume (e.g., Barclay and Warner, 1993; Eun and Sabherwal, 2003; Frijns et al., 2015), while others consider also the transaction volume of each trade (e.g., Lee and Radhakirshna, 2000).

Our third set of variables represents two controls. In particular, we use the daily log-returns on Bitstamp exchange (ret_BTSP_t) to asses whether the direction of the spot returns affects price discovery. Finally, we include the daily log-returns of the front-end contract of the Gold futures (COMEX), denoted as ret_Gold_t, serving as a proxy for the demand for financial safety in times of economic turmoil.

Table 6 reports descriptive statistics for the market quality measures that we consider in our second-stage analysis. The table shows that the spot market (Panel A) has a lower quoting and trading activity than the futures market (Panel B) over the full sample period. In particular, the daily average number of quotes is 33,532 and 56,688 for the spot and futures market, respectively. Moreover, the average traded volume is higher on the futures market (14,258) than on the spot market (9,777). For trading costs, we find that the spot market is the cheaper market. Finally, we report summary statistics for the different trading size groups. These figures reveal that the number of trades is much higher on the spot market than on the futures market. The explanation underlying this result refers to the fact that bitcoin is divisible into smaller units, while this is not possible on the futures market. Especially the number of small trades is exceptionally high on the spot market. The possibility of trading bitcoin contracts in smaller fractions potentially attracts retail investors allowing them to participate with a small investment.[15] Hence, the number of trades is higher on the spot market, while trading volume is higher on the futures market.

[Table 6 about here]

[15] The minimum unit of bitcoin is the "Satoshi", which is 0.00000001 bitcoin.

4.2 Empirical analysis

To assess the influence of the three sets of variables on the Component Share, we estimate the following equation:

$$logit\_CS_t = \beta_0 + \delta'\text{MarketQuality}_t + \gamma'\text{Uncertainty}_t + \lambda'\text{Controls}_t + \varepsilon_t, \qquad (3)$$

where $logit\_CS_t$ is the logit transformation of the spot market Component Share, which allows the mapping of the original variable, which was bounded between zero and one, to the other variables. Market Quality, Uncertainty, and Controls are the respective vectors of variables presented in Section 4.1, where we apply the natural logarithm. We further use the Variance Inflation Factor (VIF) to test for multicollinearity in Equation (3). The VIF is always below 3.54 for all our subsequent analyses, suggesting that multicollinearity is not an issue in our setting. However, we observe a relatively high correlation between spot volatility and relative trading volume (76%). Hence, we estimate Equation (3) with and without spot volatility.

We analyze the relation between Component Share and explanatory variables for two periods. First, we consider the whole sample period, which ranges from December 2017 to March 2019. Second, we look at the period from March 2018 through March 2019, which leaves out the establishment stage of the CME bitcoin futures market (futures transactions volume started very low (see Hale et al., 2018) and we avoid any liquidity issues by leaving out the first three months of trading). This reduced sample seems to be more reliable with regard to the explanatory power.

Following Frijns et al. (2015), we consider potential endogeneity issues when investigating the determinants of price discovery. In particular, we expect reverse causality between variables of market quality and CS. An improvement in price discovery may enhance several aspects of market quality. Concurrently, lower trading costs, increased liquidity, or trading volume may improve price discovery as well. Since the presence of simultaneity would produce biased estimates in an OLS framework, we employ a 2SLS estimator to capture the influence of market quality on CS.

Unreported tests reveal that the relative number of quotes, as well as the relative trading volume, are potentially endogenous.[16] We use lag one of relative number of quotes, relative trading volume, and CS, as internal instruments (see Wintoki et al., 2012; Frijns et al., 2015, for a similar procedure). Various specification statistics show that our instruments are valid and that we cannot reject the null hypothesis of exogeneity of our instruments (see Frijns et al., 2015, for similar results on diagnostic statistics). Table 7 reports the results.

[Table 7 about here]

Market quality

The results for the impact of market quality on price discovery (CS) show that the number of quotes is insignificant in all model specifications and considered time periods, indicating that there is no relation between price discovery and quoting activity. In contrast, the relative trading volume exhibits positive and significant coefficients at the 5% and 10% levels in models (1) and (2), respectively. This result indicates that an increase in trading volume on Bitstamp spot market relative to the CME bitcoin futures market is associated with an increase in price discovery on the spot market. For the whole sample period, however, the significance of relative trading volume disappears. Recall that, as discussed before, the relations between price discovery and explanatory variables may be distorted in the full sample period due to the maturing stage of CME futures.

[16] The detailed results of our pre-analysis are available upon request. A comprehensive description of the underlying intuition of the conducted tests is provided in Wintoki et al. (2012) and Frijns et al. (2015).

For the relative spreads all four specifications show negative and significant coefficients, suggesting that a decrease of trading costs in the spot relative to the futures market leads to an increase in price discovery of the spot market and vice versa. These results confirm that the cost of trading is an important determinant of where (informed) traders execute their trades and where information enters the market.

Overall, these findings are in line with the results on other asset classes, such as foreign exchange rates (see, e.g., Chen and Gau, 2010) or volatility products (see, e.g., Fernandez-Perez et al., 2018).

Uncertainty

The uncertainty variables exhibit negative coefficients that are significant for the volatility of the spot market in model (1) and for the VIX in models (1) and (2), while the lagged EPU has no (significant) effect. This implies that higher market volatility in the bitcoin spot market and higher fear in the stock market tends to increase price discovery on the bitcoin futures market.

The significant negative impact of spot market volatility on price discovery indicates that during times of high spot volatility, (informed) traders prefer to trade in the futures market. This finding could be a result of the hedging role of the bitcoin futures market when risk increases on the spot market. Chen and Gau (2010) find similar results on foreign exchange spot and futures markets, while Chakravarty et al. (2004) discover the opposite channel on stock and option markets. For the significantly negative relation between the VIX and the CS of the spot market, there is no straightforward ex-ante intuition as to why information enters bitcoin futures markets during times of high stock market volatility. Given the negative relation between VIX and bitcoin price (see, e.g., Kjaerland et al., 2018) and the interpretation of the VIX as a general fear measure, the underlying reason may also be related to hedging demand.

Once we target the whole sample period, however, these coefficients are no longer significantly related to price discovery, as before.

Controls

The control variables are all insignificant in all specifications, indicating that there is no effect of control variables on price discovery.

Additionally, we consider the influence of the trade size (small/medium/large trades) on price discovery (for definition, see Section 4.1). This analysis refers to the question of which trades have the highest price impact. Previous studies (e.g., Barclay and Warner, 1993; Chakravarty, 2001; Eun and Sabherwal, 2003) document that most information is conveyed by institutional investors, who use medium-sized orders. The so-called stealth trading hypothesis (Barclay and Warner, 1993) indicates that investors avoid to give away their information too easily by splitting large trades into smaller orders. Medium-sized orders emerge as an optimal point between trading costs and the price impact of transactions (e.g., Chakravarty, 2001).

We report the results of the different trading volume groups in Table 8 for the period March 2018 through March 2019. In line with the previous studies, the relative number of medium-sized trades is statistically significant, while the relative number of small and large trades are insignificant. In addition, the relative number of quotes reveals a negative and significant coefficient in model (2). Likewise, the Bitstamp returns turn significant in specifications (1), (2), and (4).

The results of the different trading volumes suggest that medium-sized orders are more informative than small and large trades. Hence, the more medium-sized trades occur in one market relative to the other market, the higher, on average, is the price discovery in the respective market. This finding is consistent with the stealth trading hypothesis mentioned above. Due to our data structure, however, we cannot evaluate which (informed) investors (e.g., bitcoin miners, banks, or exchanges) conduct these medium-sized trades.

[Table 8 about here]

We perform two additional tests to check the robustness of our second-stage analysis further. First, we carry out the 2SLS regressions using the price discovery results of our first-stage analysis, which were produced by the other deterministic components, i.e., constant and restricted trend (see Johansen 1995). Our results using these different specifications are qualitatively similar to those reported in Table 7. As a second robustness check, we estimate Equation (3) by adding a dummy for the halt of futures trading on CBOE (15 March 2019).[17] We do not find any change in our results.

5 Conclusion

This paper examines the evolution of bitcoin price discovery as well as the determinants of the calculated price discovery measure. Using Component Share and Information Share in our first stage of the analysis, we find that the price discovery measures are subject to time variation,

[17] The included dummy is positive and highly significant confirming the increase in the Component Share of the spot market.

suggesting that the leading market has changed over time. These findings reveal that price discovery is not limited to the spot market when considering the most liquid contract on each day. In particular, our results show a clear price leadership of the futures market in mid of 2018. On the contrary, we find evidence that the spot market is the leading market at the end of our sample. Our robustness analysis with increased time intervals shows that the information transmission between spot and futures market takes less than fifteen minutes.

In our second stage, we find strong evidence that the relative bid-ask spread negatively affects price discovery. Furthermore, we show that the relative trading volume has a positive effect on price discovery that is, however, not always statistically significant. For the relative number of quotes, we find no effect on price discovery. We further document a negative relationship between spot market volatility and price discovery, which we attribute to the hedging demand of informed investors in times of high spot market volatility. Among the control variables, we do not find an effect on price discovery. Finally, we report that medium-sized trades affect the price discovery process most, suggesting that institutional investors potentially split large trades into medium-sized trades. In conclusion, our results imply that an enhancement in market quality, such as lower trading costs and higher trading activity, has a positive causal effect on price discovery.

The bitcoin, as an emerging innovation in recent years, has received much attention due to its unique features. Despite the still existing ambiguity of the bitcoin universe, our research shows that, at least, the analysis of determinants on price discovery leads to economically reasonable results, which can also be found in other asset classes. However, the causal channel between VIX and price discovery is still unclear at this point.

Of course, comprehensive data on participating traders, and their classification into informed and uninformed traders, would allow us to even better explain the observed time variation in price discovery. For example, there is anecdotal evidence that bitcoin miners participate in the bitcoin futures market when prices move towards the mining costs. This may cause the futures market to lead the price discovery in this phase as miners potentially hedge downside risk. Unfortunately, we cannot address the underlying structure in price discovery shifts in more-depth as we have no data on the involved players.

Appendix

Appendix A: Component Share (CS) measure

Following Baillie et al. (2002) we compute the daily Component Share as

$$\gamma_{Spot,t} = \frac{\alpha_t^{Futures}}{\alpha_t^{Futures} - \alpha_t^{Spot}}, \tag{4}$$

where γ_{1t} is the Component Share of the spot market on day t. Likewise,

$$\gamma_{Futures,t} = 1 - \gamma_{Spot,t}. \tag{5}$$

The CS equation does not prevent the error-correction coefficients from being negative. Since the size, and not the sign, plays an important role in the price discovery process, we follow Cabrera et al. (2009) and restrict the factor weights to be positive. In our case of a two-market system, we define the CS as

$$CS_{1,t}^{Spot} = \gamma_1 = \frac{|\alpha_t^{Futures}|}{|\alpha_t^{Futures}|+|\alpha_t^{Spot}|} \text{ and } CS_{2,t}^{Futures} = \gamma_2 = \frac{|\alpha_t^{Spot}|}{|\alpha_t^{Futures}|+|\alpha_t^{Spot}|}, \tag{6}$$

where $CS_{1,t}^{Spot}$ is the daily Component Share for the bitcoin spot market, and $CS_{2,t}^{Futures}$ is the daily Component Share for the bitcoin futures market. The sum of the Component Shares equals one.

Appendix B: Information Share (IS) measure

Skipping the VMA representation, Hasbrouck (1995) defines $\psi\Omega\psi'$ as the variance of the common factor shocks. If we assume that two markets of interest are uncorrelated, then Ω is diagonal, and the information share IS_j of the distinct market j to the total variance is given by

$$IS_j = \frac{\psi_j^2 \Omega_{jj}}{\psi\Omega\psi'}, \tag{7}$$

where ψ_j is the contribution of the corresponding market to the total variance. Following Baillie et al. (2002), we compute the Information Share directly from the results of the VECM. The authors show that $\psi = (\psi_1 \quad \psi_2)$ is directly related to the common factor component, which means that

$$\frac{\psi_1}{\psi_2} = \frac{\gamma_1}{\gamma_2}. \tag{8}$$

Thus, we can substitute (8) into (7) and receive the contribution of the market shocks on one market to the total variance, i.e., the information share, as

$$IS_j = \frac{\gamma_j^2 \sigma_j^2}{\gamma_1^2 \sigma_1^2 + \gamma_2^2 \sigma_2^2}, \tag{9}$$

where j represents the market of interest, and σ_1^2 and σ_2^2 is the variance of the bitcoin spot and futures, respectively. If the innovations of the two markets are contemporaneously correlated, i.e., $\rho \neq 0$, Hasbrouck (1995) uses the Cholesky factorization of $\Omega = \mathrm{MM}'$ to adjust for the correlation. The Information Shares can be expressed in our bivariate market system as

$$IS_{1,t}^{Spot} = \frac{(\gamma_1 m_{11} + \gamma_2 m_{12})^2}{(\gamma_1 m_{11} + \gamma_2 m_{12})^2 + (\gamma_2 m_{22})^2} \text{ and } IS_{2,t}^{Futures} = \frac{(\gamma_2 m_{22})^2}{(\gamma_1 m_{11} + \gamma_2 m_{12})^2 + (\gamma_2 m_{22})^2}, \tag{10}$$

where $\mathrm{M} = \begin{pmatrix} m_{11} & 0 \\ m_{12} & m_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{\sigma_2(1-\rho^2)} \end{pmatrix}$, and γ_j is the contribution of each market to the total innovations. Since the calculation of the Information Shares is impacted by the order of the market price series in the Cholesky factorization, we follow Baillie et al. (2002) transposing the order of the bitcoin spot and futures markets, and take the simple average of the lower and upper bound.

Table A1

VEC model results

	(1) Obs.	(2) $alpha_{spot}$	(3) $alpha_{fut}$	(4) β_{mean}	(5) β_{median}
Panel A: Contract by contract					
BTCF8	23	-0.0892	0.0542	-0.9982	-0.9908
		(-4.767)	(3.692)	(0.121)	
BTCG8	12	-0.0705	0.0722	-1.0177	-1.0093
		(-2.025)	(2.668)	(-0.857)	
BTCH8	21	-0.0963	0.0395	-1.0970	-0.9867
		(-5.269)	(2.017)	(-0.768)	
BTCJ8	19	-0.0965	0.0569	-0.8704	-0.9647
		(-5.092)	(1.834)	(1.699)	
BTCK8	14	-0.0911	0.0836	-0.9667	-0.9694
		(-7.710)	(3.320)	(2.612)	
BTCM8	19	-0.1376	0.0142	-0.4741	-0.9550
		(-7.152)	(0.990)	(1.077)	
BTCN8	11	-0.1264	0.0171	-0.9606	-0.9889
		(-9.341)	(1.027)	(2.555)	
BTCQ8	22	-0.1067	0.0492	-0.9686	-0.9723
		(-5.267)	(2.162)	(3.999)	
BTCU8	15	-0.1414	0.0084	-0.9125	-0.9425
		(-7.820)	(0.322)	(3.918)	
BTCV8	14	-0.0461	0.0284	-0.9566	-0.9506
		(-5.812)	(2.749)	(1.085)	
BTCX8	16	-0.0549	0.0548	-0.9291	-0.9460
		(-3.985)	(2.330)	(2.721)	
BTCZ8	17	-0.0501	0.1170	-0.9538	-0.9653
		(-3.066)	(4.341)	(2.072)	
BTCF9	15	-0.0868	0.0776	-0.3182	-0.9552
		(-3.165)	(2.843)	(1.052)	
BTCG9	13	-0.0527	0.1182	-0.9752	-0.9441
		(-5.063)	(5.414)	(0,660)	
BTCH9	19	-0.0362	0.1299	-0.9700	-0.9673
		(-5.785)	(11.752)	(2.540)	
Panel B: All Data (December 18, 2017 – March 31, 2019)					
	250	-0.0858	0.0615	-0.8931	-0.9653
		(-16.944)	(10.253)	(1.943)	

This table reports the results of the VECM as presented in (1), based on one-minute sampled midquotes on CME. The VEC model is estimated every day, and the average coefficients, as well as the respective t-statistics (in parentheses), are presented for each considered contract. Additionally, we present the median of the Beta estimation. Rank of co-integration is estimated by Likelihood-Ratio test. SBIC is used to identify the daily lag length.

References

Ahking, F. W., 2002. Model mis-specification and Johansen's co-integration analysis: an application to the US money demand. Journal of Macroeconomics 24, 51-66.

Baillie, R. T., Booth, G. G., Tse, Y., Zabotina, T., 2002. Price Discovery and common factor models. Journal of Financial Markets 5, 309-321.

Baker, S. R., Bloom, N., Davis, S. J., 2013. Measuring economic policy uncertainty. Chicago Booth Paper 13-02.

Barclay, M. J., Warner, J. B., 1993. Stealth trading and volatility: Which trades move prices? Journal of Financial Economics 34, 281-305.

Baur, D. G., Dimpfl, T., 2019. Price Discovery in bitcoin Spot and Futures? Journal of Futures Markets 39, 1-15.

Boehme, R., Christin, N., Edelman, B., Moore, T., 2015. Bitcoin: Economics, Technology, and Governance. Journal of Economic Perspectives 29, 213-238.

Booth, G. G., So, R. W., Tse, Y., 1999. Price Discovery in the German Equity Index Derivatives Markets. Journal of Futures Markets 19, 619-643.

Bouri, E., Jalkh, N., Molnár, P., Rouband, D., 2017. Bitcoin for energy commodities before and after the December 2013 crash: diversifier, hedge or safe haven? Applied Economics 49, 5063-5073.

Brandvold, M., Molnár, P., Vagstad, K., Valstad, O. C. A., 2015. Price Discovery on Bitcoin exchanges. Journal of International Financial Markets, Institutions & Money 36, 18-35.

Brière, M., Oosterlinck, K., Szafarz, A., 2015. Virtual currency, tangible return: Portfolio diversification with bitcoin. Journal of Asset Management 16, 365-373.

Cabrera, J., Wang, T., Yang, J., 2009. Do Futures Lead Price Discovery in Electronic Foreign Exchange Markets? Journal of Futures Markets 29, 137-156.

Chakravarty, S., 2001. Stealth-trading: which traders' trades move stock prices? Journal of Financial Economics 61, 289-307.

Chakravarty, S., Gulen, H., Mayhew, S., 2004. Informed Trading in Stock and Option Markets. The Journal of Finance 59, 1235-1257.

Chan, K., 1992. A further Analysis of the Lead-Lag Relationship between the Cash Market and Stock Index Futures Market. Review of Financial Studies 5, 123-152.

Chen, Y.-L., Gau, Y.-F., 2010. News announcements and price discovery in foreign exchange spot and futures markets. Journal of Banking & Finance 34, 1628-1636.

Corbet, S., Lucey, B., Peat, M., Vigne, S., 2018. Bitcoin Futures – What use are they? Economics Letters 172, 23-27.

Dimpfl, T., Flad, M., Jung, R. C., 2017. Price Discovery in agricultural commodity markets in the presence of futures speculation. Journal of Commodity Markets 5, 50-62.

Dyhrberg, A. H., 2016. Hedging capabilities of bitcoin. Is it the virtual gold? Finance Research Letters 16, 139-144.

Eun, C. S., Sabherwal, S., 2003. Cross-Border Listings and Price Discovery: Evidence from U.S.-Listed Canadian Stocks. The Journal of Finance 58, 2, 549-575.

Fernandez-Perez, A., Frijns, B., Gafiatullina, I., Tourani-Rad. A., 2018. Determinants of intraday price discovery in VIX exchange traded notes. Journal of Futures Markets 38, 535-548.

Foley, S., Karlsen, J. R., Putniņš T. J., 2019. Sex, Drugs, and Bitcoin: How much Illegal Activity Is Financed through Cryptocurrencies? Review of Financial Studies 32, 1798-1853.

Fricke, C., Menkhoff, L., 2011. Does the "Bund" dominate price discovery in Euro bund futures? Examining information shares. Journal of Banking & Finance 35, 1057-1072.

Frijns, B., Gilbert, A., Tourani-Rad, A., 2015. The determinants of price discovery: Evidence from US-Canadian cross-listed shares. Journal of Banking & Finance 59, 457-468.

Gonzalo, J., Granger, C., 1995. Estimation of common long-memory components in cointegrated systems. Journal of Business and Economic Studies 13, 27-35.

Grammig, J., Melvin, M., Schlag, C., 2005. Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects. Journal of Empirical Finance 12, 139-164.

Hale, G., Krishnamurthy, A., Kudlyak, M., Shultz, P., 2018. How Futures Trading Changed Bitcoin Prices. FRBSF Economic Letter.

Hansen, H., Juselius, K., 1995. Cats in rats: co-integration analysis of time series. Estima, Evanston, Illinois.

Harris, F. H. deB., McInish, T. H., Shoesmith G. L., Wood, R. A., 1995. Cointegration, error correction, and price discovery on informationally linked security markets. Journal of Financial and Quantitative Analysis 30, 563-579.

Hasbrouck, J., 1995. One Security, Many Markets: Determining the Contributions to Price Discovery. The Journal of Finance 50, 1175-1199.

Hauptfleisch, M., Putniņš T. J., Lucey, B., 2016. Who sets the price of gold? London or New York. Journal of Futures Markets 36, 564-586.

Jin, M., Li, Y., Wang, J., Yang, Y. C., 2018. Price discovery in the Chinese gold market. Journal of Futures Markets 38, 1262-1281.

Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Oxford: Oxford University Press.

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

Kjaerland, F., Khazal, A., Krogstad, E., A., Nordstrom, F. B. G., Oust, Are, 2018. An Analysis of Bitcoin's Price Dynamics. Journal of Risk and Financial Management 11, 1-18.

Köchling, G., Müller, J., Posch, P. N., 2018. Does the introduction of futures improve the efficiency of Bitcoin? Finance Research Letters, In Press.

Kristoufek, L., 2015. What Are the Main Drivers of the Bitcoin Price? Evidence from Wavelet Coherence Analysis. PLoS ONE 10, 1-15.

Kroll, J. A., Davey, I. C., Felten, E. W., 2013. The Economics of Bitcoin Mining or, Bitcoin in the Presence of Adversaries. Working Paper, Princeton University.

Lee, C. M. C., Radhakrishna, B., 2000. Inferring investor behavior: Evidence from TORQ data. Journal of Financial Markets 3, 83-111.

Li, X., Wang, C. A., 2017. The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. Decision Support Systems 95, 49-60.

Mai, F., Shan, Z., Bai, Q., Wang, X. S., Chiang, R.H.L., 2018. How does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. Journal of Management Information Systems 35, 19-52.

Nakamoto, S., 2008. Bitcoin: A peer-to-peer electronic cash system: Retrieved from https://bitcoin.org/bitcoin.pdf.

Pagnottoni, P., Dimpfl, T., 2019. Price Discovery on Bitcoin Markets. Digital Finance, 1-23.

Putniņš T. J., 2013. What do price discovery metrics really measure? Journal of Empirical Finance 23, 68-83.

Schlusche, B., 2009. Price Formation in Spot and Futures Markets: Exchange Traded Funds vs. Index Futures. Journal of Derivatives 17, 26-40.

Shyy, G., Vijayraghavan, V., Scott-Quinn, B., 1996. A further Investigation of the Lead-Lag Relationship between the Cash Market and Stock Index Futures Market with the use of Bid/Ask-Quotes: The Case of France. Journal of Futures Markets 16, 405-420.

Theissen, E., 2012. Price Discovery in spot and futures markets: a reconsideration. The European Journal of Finance 18, 969-987.

Tse, Y., Xiang, J., Fung, J. K. W., 2006. Price Discovery in the foreign exchange futures market. Journal of Futures Markets 26, 1131-1143.

Urquhart, A., 2016. The inefficiency of Bitcoin. Economics Letters 148, 80-82.

Wang, Y., Chen, C. R., Huang, Y. S., 2014. Economic policy uncertainty and corporate investment: Evidence from China. Pacific-Basin Finance Journal 26, 227-243.

Wintoki, M. B., Linck, J. S., Netter, J. M., 2012. Endogeneity and the dynamics of internal corporate governance. Journal of Financial Economics 105, 581-606.

Yan, B., Zivot, E., 2010. A structural analysis of price discovery measures. Journal of Financial Markets 13, 1-19.

Figure 1: Five-day moving average of price discovery measures for Bitstamp spot market

This figure plots the five-day moving average of the Component Share and Information Share on the spot market. Component Share, as well as Information Share, are calculated from one-minute sampled midquotes.

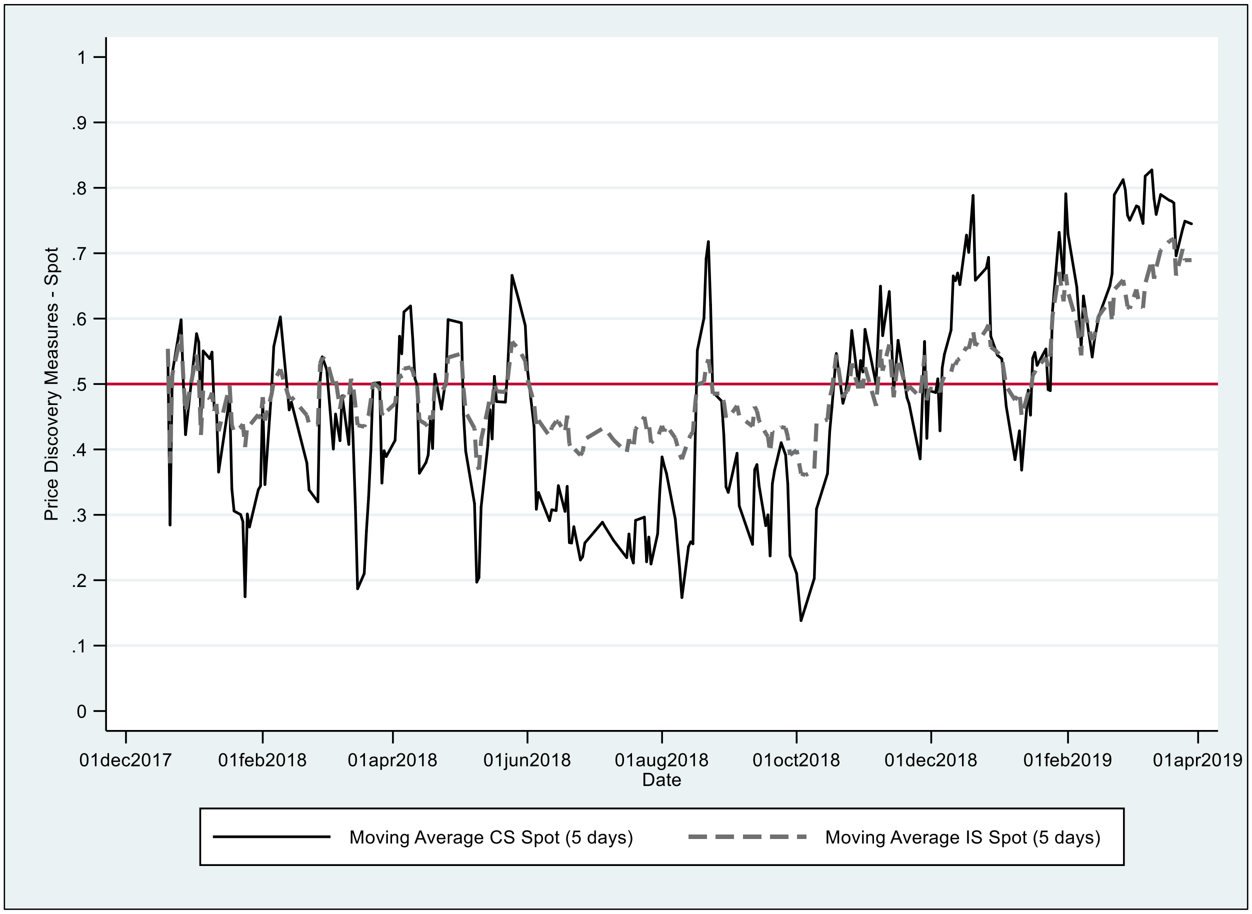

Table 1: Trading volume and average proportion of futures contracts by maturing month

(1)	(2)	(3)	(4)	(5)
RIC	Time interval most traded	Expiration	Volume MTF	Avg. Proportion
BTCF8	18Dec2017 – 25Jan2018	26Jan2018	23,457	84.80
BTCG8	26Jan2018 – 22Feb2018	23Feb2018	19,999	84.50
BTCH8	23Feb2018 – 28Mar2018*	30Mar2018*	46,090	95.27
BTCJ8	29Mar2018 – 26Apr2018	27Apr2018	62,265	94.67
BTCK8	27Apr2018 – 24May2018	25May2018	66,470	94.34
BTCM8	25May2018 – 28Jun2018	29Jun2018	57,637	94.06
BTCN8	29Jun2018 – 26Jul2018	27Jul2018	80,652	95.27
BTCQ8	27Jul2018 – 30Aug2018	31Aug2018	121,796	96.65
BTCU8	31Aug2018 – 27Sep2018	28Sep2018	46,475	94.69
BTCV8	28Sep2018 – 25Oct2018	26Oct2018	38,005	92.72
BTCX8	26Oct2018 – 29Nov2018	30Nov2018	97,089	93.95
BTCZ8	30Nov2018 – 27Dec2018	28Dec2018	56,642	94.28
BTCF9	28Dec2018 – 24Jan2019	25Jan2019	55,432	94.68
BTCG9	25Jan2019 – 21Feb2019	22Feb2019	66,737	94.31
BTCH9	22Feb2019 – 28Mar2019	29Mar2019	74,209	93.78
BTCJ9	29Mar2019 – 31Mar2019	26Apr2019#	4,237	92.43
TOTAL	327 trading days	---	815,888	93.15

This table contains several statistics on our CME futures time series. Time interval covers the days, on which the respective contract (RIC) is the most traded futures (MTF) per day. Expiration refers to the settlement date of the respective futures contract. Volume MTF is the sum of the daily volume during the provided time interval. Finally, the average proportion is defined as the trading volume of the most actively traded futures contract relative to the total trading volume in the respective time interval. * indicates that volume shifts one day earlier as the day before expiration is a holiday, while # marks that the respective contract is outside of our sample period. The sample period is from December 17, 2017 – March 31, 2019.

Table 2: Non-synchronicity and percentage of Zero Returns

Time Interval	Non-synchronous quoting (%)		Zero Returns (%)	
	Spot	Fut.	Spot	Fut.
1 minute	0.35	4.40	15.32	43.37
5 minutes	0.02	0.23	4.71	24.04
10 minutes	0.02	0.11	5.47	18.22
15 minutes	0.02	0.09	2.34	15.04

This table reports the proportion of non-synchronous quoting and percentage of zero returns of merged spot and futures time series during our sample interval from December 18, 2017 – March 31, 2019. Non-synchronous quoting is defined as the proportion of time intervals in which no quote is observed. Zero Returns (%) is the proportion of no price change. We calculate both measures for one-, five-, ten-, and fifteen-minute intervals.

Table 3: Summary statistics of Bitcoin spot and futures midquotes

Panel A: Contract by contract

(1)	(2)	(3)		(4)		(5)		(6)	
	N	Mean		Std. Dev.		NQ-Prob. (%)		ZR (%)	
		Spot	Fut.	Spot	Fut.	Spot	Fut.	Spot	Fut.
BTCF8	31,200	14,035	14,076	2,275	2,391	0.07	1.78	4.30	8.66
BTCG8	22,800	9,393	9,391	1,368	1,349	0.00	1.23	5.03	12.32
BTCH8	28,799	9,398	9,393	1,101	1,103	0.08	3.03	7.49	17.52
BTCJ8	23,999	7,765	7,769	852	855	0.02	5.75	13.59	30.82
BTCK8	24,000	8,770	8,777	579	592	0.06	3.40	13.29	28.73
BTCM8	29,998	6,925	6,924	548	560	0.03	6.30	21.98	42.61
BTCN8	22,800	6,935	6,928	709	714	0.04	3.16	15.06	41.50
BTCQ8	29,995	6,865	6,855	585	586	0.17	1.48	18.73	34.56
BTCU8	22,767	6,536	6,521	298	307	0.22	6.85	28.04	46.36
BTCV8	23,909	6,456	6,450	121	125	1.25	10.70	45.86	63.70
BTCX8	28,669	5,591	5,577	988	994	1.07	9.39	33.73	56.88
BTCZ8	21,599	3,649	3,626	274	274	0.02	1.06	15.97	39.53
BTCF9	21,590	3,727	3,703	167	165	0.30	2.44	31.70	67.49
BTCG9	22,788	3,540	3,525	178	188	1.28	2.52	39.12	80.82
BTCH9	29,962	3,883	3,878	85	88	0.61	6.20	41.18	84.76

Panel B: All data (December 19, 2017 – March 31, 2019)

	386,071	7,035	7,031	2,957	2,978	0.35	4.40	22.20	43.37

This table reports summary statistics of average midquote (mean), standard deviation (Std. Dev.), non-synchronous quoting in percent (NQ-Prob. (%)), and percentage of zero returns (ZR (%)) for each most-traded contract (Panel A) and for the whole sample (Panel B). The statistics are calculated from one-minute sampled midquotes.

Table 4: Price discovery measures for one-minute intervals

	Component Share					Information Share				
	Mean	5th Per.	Med.	95th Per.	Std. Dev.	Mean	5th Per.	Med.	95th Per.	Std. Dev.
Panel A: Contract by contract										
BTCF8	0.418	0.084	0.247	0.917	0.321	0.466	0.224	0.424	0.760	0.167
BTCG8	0.438	0.030	0.418	0.942	0.300	0.471	0.335	0.473	0.625	0.091
BTCH8	0.406	0.012	0.385	0.781	0.288	0.481	0.343	0.474	0.568	0.100
BTCJ8	0.507	0.164	0.473	0.885	0.250	0.499	0.328	0.488	0.678	0.088
BTCK8	0.442	0.062	0.401	0.784	0.272	0.472	0.316	0.474	0.643	0.106
BTCM8	0.297***	0.010	0.226	0.667	0.175	0.425***	0.269	0.426	0.561	0.071
BTCN8	0.245***	0.036	0.178	0.438	0.153	0.413***	0.314	0.421	0.483	0.064
BTCQ8	0.410	0.068	0.333	0.857	0.279	0.460**	0.360	0.449	0.572	0.072
BTCU8	0.321***	0.050	0.349	0.901	0.227	0.426***	0.331	0.409	0.569	0.066
BTCV8	0.450	0.004	0.461	0.933	0.261	0.485	0.267	0.453	0.790	0.153
BTCX8	0.511	0.043	0.456	0.940	0.299	0.497	0.220	0.482	0.487	0.145
BTCZ8	0.611	0.051	0.667	0.941	0.283	0.539	0.365	0.535	0.719	0.082
BTCF9	0.530	0.005	0.507	0.967	0.314	0.537	0.404	0.500	0.753	0.121
BTCG9	0.670***	0.197	0.716	0.995	0.213	0.608***	0.370	0.620	0.802	0.122
BTCH9	0.774***	0.367	0.868	0.944	0.165	0.682***	0.437	0.719	0.848	0.115
Panel B: All Data (December 18, 2017 – March 31, 2019)										
	0.472	0.050	0.451	0.933	0.288	0.499	0.328	0.479	0.764	0.128

Panel A reports descriptive statistics for daily price discovery measures, referring to the spot market, and estimated for each day in our sample. Panel B presents the results for the whole data set. We estimate the Component Shares (CS) and the Information Shares (IS) for one-minute time intervals. The ***/**/* are used to indicate that an estimate is significantly different from 0.50 at the 1% /5% /10% level.

Table 5: Price Discovery measures for different time intervals

	Component Share			Information Share		
	Five-MI	Ten-MI	Fifteen-MI	Five-MI	Ten-MI	Fifteen-MI
Panel A: Contract by contract						
BTCF8	0.464	0.499	0.467	0.502	0.503	0.496
	(0.367)	(0.391)	(0.439)	(0.482)	(0.493)	(0.497)
BTCG8	0.398	0.345**	0.471	0.488	0.490*	0.499
	(0.416)	(0.282)	(0.506)	(0.491)	(0.490)	(0.500)
BTCH8	0.475	0.476	0.589	0.510	0.509	0.502
	(0.496)	(0.470)	(0.613)	(0.500)	(0.499)	(0.502)
BTCJ8	0.586*	0.574	0.474	0.513	0.505	0.497
	(0.629)	(0.606)	(0.486)	(0.507)	(0.505)	(0.500)
BTCK8	0.490	0.514	0.550	0.496	0.498	0.504
	(0.522)	(0.518)	(0.565)	(0.502)	(0.501)	(0.505)
BTCM8	0.359***	0.420	0.445	0.482**	0.490	0.493
	(0.361)	(0.383)	(0.450)	(0.485)	(0.494)	(0.498)
BTCN8	0.321***	0.389	0.471	0.474**	0.481	0.488
	(0.240)	(0.356)	(0.510)	(0.478)	(0.494)	(0.500)
BTCQ8	0.475	0.488	0.496	0.492	0.497	0.500
	(0.434)	(0.455)	(0.471)	(0.497)	(0.497)	(0.500)
BTCU8	0.388	0.427	0.546	0.482*	0.491	0.499
	(0.365)	(0.452)	(0.502)	(0.487)	(0.497)	(0.499)
BTCV8	0.429	0.417	0.498	0.488	0.478	0.506
	(0.470)	(0.420)	(0.526)	(0.485)	(0.488)	(0.500)
BTCX8	0.558	0.489	0.394	0.519	0.496	0.492
	(0.473)	(0.442)	(0.374)	(0.496)	(0.496)	(0.496)
BTCZ8	0.567	0.536	0.557	0.506	0.505	0.503
	(0.616)	(0.573)	(0.568)	(0.507)	(0.501)	(0.501)
BTCF9	0.468	0.500	0.501	0.507	0.505	0.502
	(0.510)	(0.508)	(0.518)	(0.500)	(0.500)	(0.500)
BTCG9	0.688***	0.630	0.566	0.562***	0.522	0.510
	(0.649)	(0.692)	(0.571)	(0.544)	(0.525)	(0.513)
BTCH9	0.797***	0.681***	0.629**	0.614***	0.552***	0.537**
	(0.823)	(0.729)	(0.676)	(0.604)	(0.539)	(0.514)
Panel B: All data (December 18, 2017 – March 31, 2019)						
	0.501	0.497	0.516	0.510**	0.503	0.503
	(0.495)	(0.471)	(0,511)	(0.500)	(0.498)	(0.500)

Panel A of Table 5 reports average results for daily price discovery measures, referring to the spot market, and calculated for each contract in our sample from mid-quotes on CME. We also present the results for the whole data set (Panel B). We estimate the Component Shares (CS) and the Information Shares (IS) for five-, ten-, and fifteen-minute time intervals. The ***/**/* are used to indicate that an estimate is significantly different from 0.50 at the 1% /5% /10% level. Median figures are reported in parentheses.

Table 6: Summary statistics of determinants

	Mean	Median	5% quantile	95% quantile	Std. dev.
Panel A: Spot market					
Number of Quotes_t	33,532.24	35,748,00	14,814.00	48,923.00	10,978.75
Traded Volume_t	9,776.68	8,118.70	3,109.69	19,895.60	6,831.16
$\%\text{BAS}_t$	0.06103	0.0506	0.0243	0.1471	0.0377
$\text{num\_small\_trades}_t$	23,664.21	19676.00	5810.00	58392.00	18916.21
$\text{num\_medium\_trades}_t$	1,744.20	1,511.00	636.00	3,735.00	1,092.83
$\text{num\_large\_trades}_t$	266.45	197.00	50.00	705.00	248.00
Panel B: Futures market					
Number of Quotes_t	56,687.60	45,397.00	23,394.00	100,515.00	69,178.47
Traded Volume_t	14,258.26	12,315.00	3,460.00	32,500.00	9,178.10
$\%\text{BAS}_t$	0.1870	0.1410	0.0936	0.4654	0.1178
$\text{num\_small\_trades}_t$	1,687.34	1,455.00	519.00	3,650.00	1,057.58
$\text{num\_medium\_trades}_t$	455.25	378.00	35.00	1,226.00	376.68
$\text{num\_large\_trades}_t$	19.43	12.00	1.00	58.00	21.52

This table reports summary statistics of selected explanatory variables on price discovery on a daily basis for the full sample period. The considered variables of market quality are defined in Section 4.1.

Table 7: Determinants of Component Share

Variable name	*March 29, 2018 – March 31, 2019* (1) logit CS	(2) logit CS	*December 18, 2017 – March 31, 2019* (3) logit CS	(4) logit CS
ln_rel_num Quotes_t	-1.144	-0.558	-0.271	-0.060
	(-1.604)	(-1.055)	(-0.596)	(-0.165)
ln_rel_Traded Volume_t	1.554**	0.784*	0.231	-0.093
	(2.148)	(1.699)	(0.714)	(-0.561)
ln_rel_%BAS_t	-0.888**	-1.335***	-1.071***	-1.306***
	(-2.050)	(-4.451)	(-2.921)	(-4.639)
ln_$\text{vola}_{t,\text{Spot}}$	-0.896*		-0.527	
	(-1.859)		(-1.355)	
ret_VIX_t	-0.032**	-0.030**	0.003	0.002
	(-2.096)	(-2.013)	(0.228)	(0.139)
ln_EPU_{t-2}	-0.340	-0.303	-0.208	-0.203
	(-1.463)	(-1.322)	(-1.007)	(-0.979)
ret_Bitstamp_t	-0.045	-0.047	-0.030	-0.028
	(-1.425)	(-1.401)	(-1.165)	(-1.081)
ret_Gold_t	0.033	0.045	-0.003	0.012
	(0.181)	(0.243)	(-0.018)	(0.073)
Constant	-2.226	0.011	-2.303*	-0.853
	(-1.523)	(0.010)	(-1.771)	(-0.867)
Observations	194	194	250	250
Adj_R-squared	0.0880	0.105	0.0723	0.0709
Hansen's J test	0.763	0.570	0.864	0.814
Wooldrige's score test	0.291	0.551	0.516	0.721

This table reports results for Equation (3) where we assess the relationship between various explanatory variables and the logit transformation of Component Share that refers to the spot market. The model is estimated by 2SLS using robust standard errors, where the relative number of quotes and the relative traded volume are treated as endogenous and the remaining variables as exogenous. We use lag one as instruments. Robust t-figures are reported in parentheses. The ***/**/* indicate that an estimate is statistically significant at the 1% /5% /10% level.

Table 8: Determinants of Price Discovery with decomposed relative trading volume

	March 29, 2018 – March 31, 2019			
	logit CS	logit CS	logit CS	logit CS
Variable name	(1)	(2)	(3)	(4)
$ln\_rel\_num\ Quotes_t$	0.876	-0.836*	0.119	-0.003
	(0.926)	(-1.695)	(0.180)	(-0.002)
$ln\_rel\_num\_small\_trades_t$	-1.102			-0.605
	(-1.467)			(-0.821)
$ln\_rel\_num\_medium\_trades_t$		1.072***		1.180***
		(3.171)		(3.272)
$ln\_rel\_num\_large\_trades_t$			-0.238	-0.390
			(-0.720)	(-1.166)
$ln\_rel\_\%BAS_t$	-1.089***	-0.923**	-1.200***	-0.624*
	(-3.402)	(-2.536)	(-3.719)	(-1.769)
$ln\_vola_{t,spot}$	0.001	-0.709**	0.261	-0.106
	(0.005)	(-2.155)	(0.411)	(-0.164)
ret_VIX_t	-0.017	-0.029**	-0.028**	-0.026
	(-0.936)	(-2.023)	(-1.982)	(-1.539)
ln_EPU_{t-2}	-0.205	-0.335	-0.284	-0.363
	(-0.896)	(-1.465)	(-1.264)	(-1.508)
$ret\_Bitstamp_t$	-0.055*	-0.053	-0.062*	-0.066*
	(-1.647)	(-1.641)	(-1.783)	(-1.766)
ret_Gold_t	0.165	0.020	0.102	0.029
	(0.918)	(0.115)	(0.563)	(0.167)
Intercept	2.365	-3.729**	1.314	1.401
	(0.886)	(-2.244)	(0.380)	(0.310)
Observations	194	194	192	192
Adj. R-squared	0.0689	0.129	0.103	0.150
Hansen's J test	0.643	0.901	0.454	0.665
Wooldrige's score test	0.509	0.386	0.789	0.723

This table reports results from 2SLS regressions using the decomposed trading sizes as explanatory variables. The dependent variable refers to the logit transformation of the Component Share of the spot market. Relative number of small, medium, and large trades, as well as relative number of quotes, are treated as endogenous variables. We use lag one as instruments. Results are reported for the sample period from March 2018 through March 2019. Robust t-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.