Frankfurt School Blockchain Center

DECEMBER 2020

FSBC Working Paper

Using On-Chain and Market Metrics to Analyze the Value of Crypto Assets

Alexander Stober, Philipp Sandner

This paper sheds light on metrics that are correlated with the price of crypto assets. We conduct a literature analysis to derive such metrics and apply them to the market price of selected ERC-20 tokens. We find that monthly active addresses, as an approximation of a token's user base, and the Bitcoin price, as a representation of the market, share a positive relationship, while monthly active addresses of stablecoins share a negative relationship with the price of the examined token.

Ethereum and the ERC-20 standard

Ethereum, the second-largest blockchain by market capitalization, was created as an open-source blockchain that allowed developers to deploy and run their own program codes with a Turing-complete computer language on the blockchain to create a platform of decentralized applications (Buterin, 2013, p. 13). These program codes are also called smart contracts.

In November 2015, the ERC-20 standard was proposed to establish a common list of rules for tokens to run on the Ethereum ecosystem. This standard facilitated developers to program how new tokens behave on the blockchain, with smart contracts and with other tokens (Somin, Gordon, & Altshuler, 2018, p. 441). This allowed smart contracts to operate similar to conventional crypto currencies (Bheemaiah & Collomb, 2018, p. 58).[1] Soon after the

Frankfurt School Blockchain Center
www.fs-blockchain.de
contact@fs-blockchain.de

Follow us
www.twitter.com/fsblockchain
www.facebook.de/fsblockchain

Frankfurt School of Finance & Management gGmbH
Adickesallee 32-34
60322 Frankfurt am Main
Germany

inception of the ERC-20 standard, these tokens became widely adopted in the venture capital market (Ante, Sandner, & Fiedler, 2018, p. 14). On the one hand, developers could extend systems that already implemented this standard to support further ERC-20 compliant tokens with little effort. On the other hand, companies utilized this standard to issue their tokens on Ethereum's existing ecosystem "inducing the trade of these crypto-coins to an exponential degree" (Somin, Gordon, & Altshuler, 2018, p. 441). As of October 2, 2020, Etherscan listed more than 300,000 ERC-20 token contracts deployed on Ethereum's main network (Etherscan, 2020).

The importance of analyzing ERC-20 tokens

By examining the blockchain-related academic literature, we see two shortcomings. First, while the analyzed literature mostly studied Bitcoin metrics, hardly any analysis exists on the time series of ERC-20 tokens. The relevance of these tokens is given, as they cover a wide diversity of economic purposes and became an integral part in venture financing (Ante, Sandner, & Fiedler, 2018, p. 15). Second, the examined literature pays little attention to the publicly available transactional data and their relationship to the market price of crypto assets. Considering that blockchain data can be sourced with unquestionable accuracy, this opportunity paves the way to unparalleled insights.

Literature review on on-chain & market metrics

User base. The idea of studying the user base of a crypto asset is based on network effects. The rationale is that as the number of users of a communication service inclines, so will the utility for the individual user (Tucker, 2008, p. 2025). Based on this effect, Metcalfe stated that the utility of a two-way telecommunication network could be expressed by the nodes of the network (Shapiro & Varian, 1999, p. 184). For N linked devices, each device would be connected to $N-1$ other devices. Hence, the number of unique connections, and according to Metcalfe, also the value of the network V, equals $N \times (N-1)$.[2] Metcalfe's formulation would later become known as Metcalfe's Law (Metcalfe, 2013, p. 27).[3] In the literature, Metcalfe's Law has been applied to a few crypto assets. Alabi (2017) valued Bitcoin, Ethereum (excluding ERC-20 tokens), and Dash via Metcalfe's Law. He selected daily

unique addresses to approximate the user base and found that the unique addresses fit the price development and that large deviations may imply value bubbles (Alabi, 2017, p. 13). Metcalfe's Law was also applied to study Bitcoin by Wheatley et al. (2018) and the 50th largest ERC-20 tokens according to their market capitalization by Lehnberg (2018). In both cases, it was found that an exponent smaller than two better approximates the price of the tokens, and in the latter case that using monthly active addresses as a valuation metric performed better than using weekly and daily unique addresses. Thus, we expect a positive relationship between the token price and unique addresses as an approximation for the user base.

Stablecoins. Stablecoins may share a relationship with other crypto assets because of their design to serve as a stable crypto currency. They reduce the price volatility of the crypto asset market and are commonly pegged to fiat currencies like the US dollar. Stablecoins are widely utilized by exchange platforms to price crypto assets in USD without maintaining a USD bank account (Wei, 2018, p. 19). Their relative price stability paves the way for two use-cases.[4] The first use case is to act as a medium of exchange. The largest pegged coin by market capitalization as of October 2nd, 2020 Tether (USDT), is often utilized to purchase further crypto assets. The second use case is that stablecoins serve as a store of value, as the purchasing power of these pegged coins preserves the value (Blockchain, 2018, p. 13). Regular crypto assets are, due to their high volatility, less suitable to serve this function. This volatility risk is mainly present in ICOs. Token issuer will less likely convert their raised funds into fiat currencies because of the high level of friction for converting their funds. Thus, investors, expecting that the value of their crypto investment declines, may utilize stablecoins as a safe haven (Blockchain, 2018, p. 13). We expect a negative relationship between the token price and the user base of stablecoins due to the findings of Baur & McDermott (2010). They examined whether gold is used as a safe haven for major stock markets. In this comparison, gold would replace stablecoins as a store of value, while stocks would replace other crypto assets. The researchers found that gold is used as a store of value in the European and US market.[5]

Exchange platform transactions. Exchange platform transactions may be insightful for the following two reasons. Exchange platforms provide liquidity within the crypto market. Token holders willing to sell their tokens can expect to receive a fair price on these platforms due to lower bid-ask

spreads. Hence, they are motivated to move their assets on exchanges in case they have selling intentions. In contrast, token owners are incentivized to withdraw their assets from these platforms as soon as possible due to the risk of security breaches and the supposable total loss (Polasik, Piotrowaska, Wisniewski, Kotkowski, & Lightfoot, 2015, p. 15). Thus, exchange transactions could be interpreted as crypto assets attractiveness signals. If tokens are transferred to exchange platforms, we assume an "available-for-sale" signal and if tokens are received from exchange platforms we assume an investment signal. There are papers on attractiveness signals like Ciaian et al. (2016). They argued that the Wikipedia page views of Bitcoin approximated the public attractiveness of crypto currency and examined its relationship on the Bitcoin price development on two periods. They found that the short-term relationship has been stronger in the first period than in the last period, but that the views do not have an impact in the long-term (Ciaian et al., 2016, p. 1813). Li & Chong (2017) approximated the investment attractiveness of Bitcoin with Google search queries and Twitter tweets on two periods. They found that Google queries show a positive, statistically significant relationship with the Bitcoin price in the short term, while this was not the case for Twitter tweets (Li & Chong, 2017, p. 56). On this basis, we expect a positive relationship between token price and the amount of token received from central exchange platforms and a negative relationship between the token price and the amount of token moved to central exchange platforms.

Market factor. We use the price development of Bitcoin as a market factor approximation due to (i) Bitcoin's superior role in the crypto asset market, (ii) the utilization of Bitcoin as a trading partner rather than fiat currencies, especially at the inception of the asset (Adhami, Giudici, & Martinazzi, 2018, p. 71), (iii) similar price development patterns. These patterns could imply that either the price of Bitcoin affects the price of other crypto assets, or both prices are driven by the same external factors. Based on these observations, Ciaian et al. (2018) empirically examined the interdependencies among Bitcoin and the 16 largest alternative crypto currencies. They found a Bitcoin-crypto currencies price relationship that was in the short-term significantly stronger than in the long-term. Considering these aspects, an examination of the price development of Bitcoin on the ERC-20 tokens might be fruitful.

Thus, we expect a positive relationship between the token price and the Bitcoin price.

Variables

Table 1 provides an overview of the variables used to operationalize the factors. To mitigate issues of heteroscedasticity we apply the logarithmic transformation.

Table 1

Definition of variables used in the model

Variable	Description
Log (market price)	The natural logarithm of the daily closing price of the selected ERC-20 token. The daily closing price of an ERC-20 token, denominated in USD, is sourced from the worldwide token database TOKENBASE (2019) and calculated by using a volume-weighted formula that considers all exchange platforms on which the token of interest is listed.
Log (monthly active addresses)	The natural logarithm of the distinct number of daily active addresses, accumulated over the last 30 days. Due to Ethereum's pseudo-anonymous structure, the number of actual network users cannot be precisely quantified. As Alabi (2017, p. 3) pointed out, it can be assumed that the ratio between the actual user base and the unique addresses is roughly constant, given a large user base.[6]
Log (bitcoin price)	The natural logarithm of the daily closing price of Bitcoin. The daily closing price of Bitcoin is demonimated in USD.
Tokens to exchange	The number of daily tokens sent to selected exchange addresses divided by all tokens exchanged on a day of a selected ERC-20 token.[7]
Tokens from exchange	The number of daily tokens received from selected exchange addresses divided by all tokens exchanged on a day of a selected ERC-20 token.
Monthly active addresses of stablecoins	The natural logarithm of the distinct number of daily active addresses of active stablecoins accumulated over the last 30 days. We analyzed the transactions of all stablecoins on the Ethereum environment that had a market capitalization >0 as of July 15, 2019.

Data acquisition

The ERC-20 tokens to be analyzed are selected based on the following criteria. First, the top tokens according to their market capitalization as of July 15, 2019, have been shortlisted (CoinMarketCap, 2019). Second, we excluded pegged payment tokens, since their value is predefined and recently launched

tokens, due to short time-series. Third, we also excluded tokens that migrated to their mainnet, respectively are used on several public blockchains, since not all of their transactions happened on Ethereum's mainnet.[8] The on-chain data of the Ethereum blockchain is sourced from a relational database, provided by eth.events (2019). We used the contract addresses of the tokens to filter the relevant transactions. The same process was done for the group of stablecoins.

Methodology

To avoid spurious results when analyzing the time series of each token, it is essential to test the characteristics of time processes which may exhibit interdependency and non-stationarity, meaning that its mean, variance, and covariance are not constant over time. Thus, an OLS regression may not be the appropriate choice.

ARDL & ECM. The ARDL model is utilized to form relationships among variables in a single equation. This model consists of the lagged values of the dependent variable as well as the lagged and the current values of the independent variables (Pesaran & Shin, 1999, p. 372). The ARDL model is commonly applied because the cointegration of non-stationary variables is on par with the error correction process. This process can be derived from the ARDL model with a linear transformation (Nkoro & Uko, 2016, p. 79). In this case, the ARDL and ECM allow unraveling short-term dynamics and long-term relationships (Kripfganz & Schneider, 2018, p. 3). If the non-stationary variables are not cointegrated, which will be tested with the bounds test, then there exists no long-term equilibrium, and the ARDL is applicable. In this case, one can only derive short-term relationships. The ARDL model for one dependent and one independent variable is determined as:

$$y_t = c_0 + c_1 t + \sum_{i=1}^{p} \emptyset_i y_{t-i} + \sum_{i=0}^{q} \beta_i x_{t-i} + \epsilon_t,$$

where y is the dependent variable at time t, x denotes the independent variables, c_0 represents the intercept term, c_1 is the time trend, p is the number of lagged dependent variables, q is the number of lagged independent variables, and ϵ the disturbance term. A long-run relationship between the variables

exists if the variables that are stationary in level form and first difference exhibit cointegration. In this case, the ECM is applied. The formula for one dependent and one independent variable is determined as:

$$\Delta y_t = c_0 + c_1 t + \alpha(y_{t-1} - \theta x_t) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \psi'_{xi} \Delta x_{t-i} + \epsilon_t,$$

where α denominates the speed of adjustment coefficient, θ the long-term coefficients, ψ the short-term coefficients, Δ the difference between the current and the last period, p is the number of lagged dependent variable, q the number of lagged independent variables, and ϵ the disturbance term. If there is more than one cointegration among the examined variables, then the coefficients of the single equation of the ARDL/ECM may be misspecified.

The short-term coefficients explain the variation that is not caused by the long-term equilibrium. The long-term coefficients show the equilibrium effects of the predictor variable on the predicted variable. The adjustment coefficient measures the intensity of a deviation from the equilibrium relationship on the dependent variable (Kripfganz & Schneider, 2018, p. 15). A significant short-term coefficient exhibits that the associated variable affects the dependent variable significantly in the short-term. A time process with significant short- and long-term coefficients points to stronger effects on the dependent variable than only significant short-term coefficients (Ciaian et al., 2018, p. 187). Advantages of the ARDL approach are that (i) endogeneity is a subordinated issue since all variables are assumed to be endogenous, (ii) the model can differentiate between dependent and independent variables when there is a single long-term relationship, (iii) it can identify cointegrating vectors, and (iv) the ECM can be derived from the ARDL (Nkoro & Uko, 2016, p. 79). To test if the ECM is applicable, the time series data needs to be examined beforehand as this approach is only applicable for variables that are stationary at the level form or first difference. Hence, this thesis will first examine at which level the time process becomes stationary. Subsequently, the next step will be to analyze whether the variables that are non-stationary at the level form but stationary at first difference follow a long-term relationship. Based on these results, the tokens are examined in their short- and long-term properties.

Unit root test. According to Greene (2007, p. 754), the time series of exchange rates often exhibit non-stationarity. Ignoring to test and adapting for unit root, which assumes that the time series will not return to a deterministic path in the long-run, could cause spurious regressions (Granger & Newbold, 1974, p. 111). Hence, the time process of every token and every variable needs to be analyzed. This will be employed with the Augmented Dickey-Fuller (ADF) test for unit-root. The null hypothesis is that the time process is non-stationary, respectively exhibits a unit root. The ADF formula is:

$$\Delta y_t = \alpha + \delta t + \beta y_{t-1} + \sum_{i=1}^{k} \gamma_i \, \Delta y_{t-i} + \epsilon_t,$$

where y is the examined variable, Δy is the first difference of y, α is a constant, δ is a time trend coefficient, β is the process root coefficient, k is the number of specified lags, which will be selected according to the Akaike information criterion (AIC),[9] and ϵ is the residual term. The ADF formula considers testing for a unit root with (i) a random walk ($\alpha, \delta = 0$), (ii) a random walk with a constant ($\delta = 0$), and (iii) a random walk with a constant and a time trend. In the case that variables are non-stationary at the level form, they could be stationary in the first difference. If this is true, they are integrated of order one, respectively I (1) (Shrestha & Bhatta, 2018, p. 75).

The results of the unit root test at the level form can conclude that all variables are stationary, all variables are non-stationary, or some variables are stationary while others are non-stationary. In the first scenario, further analysis could be conducted with an unrestricted vector autoregression (VAR), or OLS estimation method (Shrestha & Bhatta, 2018, p. 74). For the other two cases, the time series needs to be examined whether they are stationary at further difference levels. In general, all variables can be differenced to the point where they become stationary and can then be examined with the OLS regression. However, the differences will then only display the short-term change in the time series and cause a loss of important long-term information (Nkoro & Uko, 2016, p. 68). Thus, to preserve valuable long-term information, the properties of time series that are stationary at the level or the first difference form will be examined on cointegration.

Cointegration test. The econometric concept of cointegration assumes that the combination of non-stationary variables converges over time (Shrestha & Bhatta, 2018, p. 77). Firstly formalized by Granger (1981), non-stationary time processes, which individually drift away from the mean, can be combined to ensure that the state of equilibrium is stable over time. If no cointegration is found, differenced variables must be utilized. Hence, if applicable, cointegration allows a solid "statistical and economic basis for empirical error correction model" (Nkoro & Uko, 2016, p. 75). Cointegration can be examined with the ARDL bounds test, proposed by Pesaran, Shin, & Smith (2001). The ARDL/ECM results become invalid when time series data is used that is stationary at a difference higher than one.[10] Hence, it is crucial to know the order of integration in advance (Nkoro & Uko, 2016, p. 79). To perform the bounds test, the absence of a level relationship is tested. This test is done by the null hypotheses that the adjustment and long-term coefficient are not significantly different from zero, which for two variables is formalized as follows:

$$\Delta y_t = c_0 + \sum_{i=1}^{k} \alpha_i \Delta y_{t-1} + \sum_{i=1}^{k} \alpha_2 \, \Delta x_{t-i} + \delta_1 y_{t-1} + \delta_2 x_{t-1} + \epsilon_{1t}$$

where y is the independent variable, x is the dependent variable, c_0 is the constant, Δ is the difference operator, α is the short-term coefficient, δ is the long-term coefficient, k is the number of lags, and ϵ is the error term. The hypothesis is that the long-term coefficients are zero, which is formalized as $H_0: \delta_1 = \delta_1 = 0$. The alternative hypothesis is defined as $H_1: \delta_1 \neq \delta_1 \neq 0)$ (Nkoro & Uko, 2016, p. 80). Possible outcomes are (i) the F-statistic is larger than the upper limit, meaning that the null hypothesis can be rejected and there is a cointegrated relationship, (ii) the F-statistic is lower than the lower limit, meaning that the null hypothesis cannot be rejected and that there is no cointegrated relationship, and (iii) the F-statistic falls between the lower and upper limit, meaning that the results are inconclusive (Ciaian et al., 2018, p. 187).

Regression analysis

First, the augmented Dickey-Fuller (ADF) test needs to be performed.[11] If the variable of interest is stationary at the level form, the ADF is not performed on the first difference. Among all ERC-20 tokens, the metrics associated with exchange addresses were always stationary at the level form. This was expected as the exchange numbers have been set into relation to the total amount of tokens transferred the same day. Hence, the number ranges between 0 and 1. The results for the other variables differ. In the case that no constant and no trend variables were used, the time series of all variables are non-stationary at level form. For the cases that a constant, or a constant and trend was applied, the examined time series became in some cases stationary at the level form. The outcome of the ADF test unveiled that all variables are either stationary at the level form or the first difference and thus, meet the requirement for the ARDL cointegration test.

Next, the bounds test is applied to examine whether the time series of the token exhibits a cointegration. If this is the case, we considered the short- and long-term estimators. If there was no cointegration, we did not further examine the corresponding token in this paper. We assumed a cointegrating relationship in case that the upper bound is exceeded at the 10% level by the F-, and t-statistic (Table 2). The cointegration test unveiled that the variables of the ERC-20 token LINK, BAT, ZRX, KCS, AE, and SNT exhibited a long-term relationship. Hence, next to the short-term effects, the long-term relationship could be drawn.

Table 3 shows the estimation results of the six tokens that exhibited cointegration. The table is separated into the categories *long-term estimators*, *short-term adjustments*, and *diagnostics*. The interpretation of the long-term estimators is that for the monthly active addresses of the LINK token, for example, a 1% increase in this metric is causing a 0.634% increase in LINK's market price in the long-run equilibrium, everything else being constant.

Table 2

Bounds test results of the selected ERC-20 tokens

ERC-20 token	F-Statistic	t-Statistic	Cointegration relationship
Binance Coin (BNB)	2.56	-3.04	No
ChainLink Token (LINK)	9.16***	-5.10***	Yes
Maker (MKR)	2.71	-3.17	No
Basic Attention Token (BAT)	4.34**	-4.20**	Yes
HuobiToken (HT)	0.63	-1.51	No
0x (ZRX)	3.38*	-3.99*	Yes
Aurora (AOA)	3.58*	-3.70	No
KuCoin Shares (KCS)	4.33**	-4.38**	Yes
Zilliqa (ZIL)	1.80	-1.04	No
EDUCare (EKT)	1.47	-1.63	No
Mixin (XIN)	0.55	-1.54	No
IOSToken (IOST)	3.21	-1.11	No
Aeternity (AE)	4.32**	-3.88*	Yes
Waltonchain (WTC)	3.08	-3.41	No
Enjin Coin (ENJ)	2.06	-3.06	No
StatusNetwork (SNT)	6.70***	-5.50***	Yes
Clipper Coin Capital (CCCX)	2.48	-2.55	No
Aelf (ELF)	2.17	-2.56	No
Golem (GNT)	3.24	-2.11	No

*The T- and f-statistic columns are decision criteria on whether to reject the joint hypothesis and rule that a cointegrating relationship is present. The critical values of the I(0) and I(1) bounds are not displayed for reasons of simple illustration. Significance levels: * 0.05 < p ≤ 0.10; ** 0.01 < p ≤ 0.05; *** p ≤ 0.01.*

Table 3

Estimation results of tokens exhibiting cointegration

	Log (market price)					
ERC-20 token	**LINK**	**BAT**	**ZRX**	**KCS**	**AE**	**SNT**
Long-term estimation						
Log (monthly active addresses)	0.634*** (0.080)	0.365*** (0.094)	0.352*** (0.061)	-0.054 (0.073)	0.520** (0.219)	0.746*** (0.103)
Tokens-to-exchange	-1.457** (0.557)	0.512 (0.706)	-0.117 (0.164)	-0.841** (0.388)	1.314 (0.834)	0.610** (0.297)
Tokens-from-exchange	0.906** (0.436)	1.051* (0.585)	0.595** (0.276)	-0.064 (0.312)	-0.601 (0.587)	0.724 (0.505)
Log (BTC market price)	0.252 (0.161)	0.526*** (0.169)	0.634*** (0.050)	1.531*** (0.271)	0.994*** (0.280)	0.560*** (0.128)
Log (monthly active addresses of stablecoins)	0.017 (0.047)	-0.191*** (0.055)	-0.367*** (0.017)	-0.199*** (0.070)	-0.346*** (0.095)	-0.274*** (0.034)
Short-term adjustment						
Δ Log (market price)$_t$	-0.119*** (0.040)		-0.085* (0.044)	-0.183*** (0.040)		
Δ Log (monthly active addresses)$_t$	1.036*** (0.117)	0.360*** (0.130)	0.169*** (0.054)	0.040*** (0.012)	0.182** (0.078)	0.339*** (0.085)
Δ Log (monthly active addresses)$_{t-1}$		0.367*** (0.140)		-0.025* (0.013)	0.170** (0.079)	
Δ Tokens-to-exchange$_t$	0.119*** (0.034)				-0.040* (0.023)	
Δ Tokens-to-exchange$_{t-1}$	0.067** (0.028)				-0.045** (0.021)	
Δ Tokens-from-exchange$_t$	-0.092*** (0.028)		-0.073** (0.028)			-0.050** (0.024)
Δ Tokens-from-exchange$_{t-1}$	-0.075*** (0.023)		-0.073*** (0.024)			
Δ Log (BTC market price)$_t$	0.972*** (0.068)	1.003*** (0.055)	0.874*** (0.062)	0.988*** (0.069)	1.087*** (0.052)	0.993*** (0.053)
Δ Log (BTC market price)$_{t-1}$	0.129* (0.078)	-0.125** (0.056)		-0.200** (0.079)		
Δ Log (stablecoin monthly active addresses)$_t$			-0.347** (0.159)			
Δ Log (stablecoin monthly active addresses)$_{t-1}$		-0.207*** (0.073)				
Constant	-0.690*** (0.224)	-0.414*** (0.140)	-0.779*** (0.235)	-0.344*** (0.115)	-0.296* (0.172)	-0.786*** (0.165)
Diagnostics						
N	571	571	315	571	571	571
F value	20.46***	22.74***	21.00***	21.55***	35.99***	25.42***
Adjustment factor	-0.081	-0.049	-0.127	-0.033	-0.029	-0.062
R2	0.400	0.440	0.530	0.384	0.493	0.467
Adjusted R2	0.381	0.420	0.505	0.366	0.479	0.449
Durbin-Watson test	2.012	2.092	1.941	1.980	1.919	1.983

*Significance levels: * 0.05 < p ≤ 0.10; ** 0.01 < p ≤ 0.05; *** p ≤ 0.01. For the short-term adjustment: Empty cells are not significantly different from zero.*

The coefficients of the logarithmically transformed monthly active addresses have been significant at the 1% level for the LINK, BAT, ZRX, and SNT token, while AE is significant at the 5% level. This metric was only for the KCS token not statistically different from zero. Furthermore, a clear trend for the tokens transferred to exchanges was not observable. While this metric exhibited significance for LINK, KCS, and SNT, the algebraic sign changed for the latter token. Considering the tokens that are received from exchange platforms, only LINK, and ZRX displayed significance, while BAT exhibited a weak significance. All of them show a positive coefficient. The other tokens were insignificant for this metric. The logarithmic transformation of the Bitcoin as a representation of the market unveiled a positive and strong significance for all token, except for LINK. The short-term adjustment effects show the first difference movements of each metric. The interpretation is that the coefficients measure the contemporaneous ceteris paribus effect of a change in the first difference on the dependent variable. The last day of LINK's market price, for example, was significantly negative related to LINK's market price. Besides this token, also the last day market price of KCS shared a significant negative relationship to its current market price. This relationship was weakly significant for ZRX, and not significant for the other assets.

The first difference for the monthly active addresses exhibited a positive relationship among all cointegrated tokens. This was also the case for the first difference of Bitcoin's market price. Tokens that were sent to exchanges share a positive relationship in the case of LINK and a negative relationship in the case of AE to their market price. Tokens that were received from exchange addresses share a negative relationship with its market price in the case of LINK, ZRX, and SNT. Lastly, the difference in the monthly active addresses of stablecoins only impacted the ZRX token in the short run for the contemporaneous day.

The diagnostic section of Table 3 includes the adjustment factor for the error correction term of each token, the number of observations, the model fit, and the Durbin-Watson test. Of particular interest is the adjustment factor of each token, as it contains meaningful information about the model. The error correction coefficient of a stable model must lie between -1 and 0 to show model stability. If the term falls out of these boundaries, there would be an over-

correction towards the equilibrium, making the model unstable. According to the crypto assets exhibiting cointegration, the adjustment factor of all coefficients was negative and significant. This was expected, as the bounds test found cointegration for these tokens and the significance of this coefficient is based on the bounds test findings. Looking at the magnitude of the error correction term of the ZRX token, for example, this number implied that 12.7% of the disequilibrium was adjusted within one day. This disequilibrium was caused by short-run shocks between the tokens market price and the applied metrics. Another meaningful control was the test for serial correlation employed with the Durbin-Watson statistic. This statistic examined the residuals of each token and implied, based on a scale from 0 to 4, whether autocorrelation was present, and if yes, if a positive or a negative autocorrelation was observed.[12] According to the estimation results of Table 3, none of the six ERC-20 tokens exhibited autocorrelation.

Interpretation of the results

User base. The long-term estimation mostly confirms that there is a positive, statistically significant relationship for LINK, BAT, ZRX, AE, and SNT. There is no case of a significant, negative coefficient. A reason why the coefficients of these assets are not significant for all tokens could be that they are predominantly exchanged on central platforms due to their utility purpose. Trades on these platforms cannot be traced, and the results become distorted. A more detailed elaboration of the on-chain data may give further insights.

Stablecoins. An examination of the long-term estimation shows that, except for LINK, all tokens exhibit negative, statistically significant coefficients. The monthly active addresses of stablecoins are also statistically significant for BAT and ZRX in the short-run.

Exchange platforms. The long-term examination of transactions to exchange addresses unveils that there is a significant relationship between this metric and the market price in the case of LINK, KCS, and SNT, however, with different algebraic signs. These inconclusive results are further blurred by the short-term adjustments as this metric displays significance only in the case of LINK, with positive and negative coefficients. Blurred results are also

provided with the metric concerning the tokens that are received from exchange addresses. LINK and ZRX show a significant positive relationship, while BAT exhibits a positive, weakly significant relationship. Based on these findings, there might be more to discover for the tokens that are associated with exchange transactions, considering that there are days where nearly all token transfers are associated with exchange addresses. However, the results of the ARDL/ECM are inconclusive in the current form. Thus, we conclude that these metrics have no explanatory power.

Market factor. The long- and short-run estimations confirmed that the Bitcoin price as the market factor has a positive, statistically significant relationship with the market price of all examined tokens at the 1% level. The LINK token is the exception for the long-run estimator, however, shows significance for the first difference. In the case of LINK, BAT, KCS, the Bitcoin price has been significantly different from zero in the short-run.

Conclusion

We examined the relationship of on-chain and market metrics on the price development of six ERC-20 tokens. This paper contributes by shedding light on the price drivers of the ERC-20 environment as previous literature primarily highlighted metrics that affected the Bitcoin price. The ecosystem of Ethereum is of particular interest due to its ability to represent a variety of tokens with a variety of purposes. Out of this environment, many projects grew and started their own mainnet, emphasizing the importance of comprehending these tokens. Based on the publicly available blockchain transactions and the market data, we tested the metrics that have been derived from our literature review. The metrics are either directly determined from the literature or derived from related investment classes. The examined metrics are monthly active addresses of the token of interest and stablecoins as a whole, the daily transactions from and to exchange addresses, as well as the Bitcoin price as the market factor. We applied the ARDL/EC model to address the issues of non-stationary time series. The examination on the tokens unveiled that out of the analyzed tokens, six exhibited cointegration.

Our results suggest that (i) monthly active addresses share a long-run relationship with the examined tokens that (ii) monthly active addresses of stablecoin share in five cases a negative long-run relationship with the examined token, that the chosen metrics (iii) tokens-to-exchange and (iv) tokens-from-exchange do predominantly not impact the market price of the token of interest, and that (v) Bitcoin as a market approximation has a positive relationship with all examined token, except for LINK in the long-term. This analysis is a humble, initial approach to unveil relationships among on-chain and market metrics on the price development of crypto assets.

Limitations. There are some arising limitations considering the utilized data. As the user base is not known, it has to be approximated with unique addresses. This complicates the comparability among the selected tokens. Aggravating this situation is the inability to access the trades that happen on exchange platforms publicly. Transactions to and from these addresses can be traced, but the activities on these platforms are veiled.

Furthermore, there are two arising issues with stablecoins. First, not all pegged coins are based on ERC-20 tokens. As these coins are not depicted on the Ethereum Blockchain, this paper was not able to acquire related data. Second, this paper tracked the stablecoins that had a market capitalization of more than zero at the 15 July 2019. This approach accounted for the most active tokens. However, tokens that have been active and became worthless before the date of interest are not considered.

From the perspective of the exchange addresses, there are also shortcomings. Since it is not possible to comprehend which exchange uses which address from Ethereum's mainnet, this paper relies upon the completeness of the exchange list that was gathered from Etherscan (2020). This list could be incomplete, in which case, the metrics associated with exchange addresses became distorted. Besides, the limitations of Ethereum's ecosystem have to be highlighted. Since all ERC-20 tokens are based on one blockchain, advantageous as well as disadvantageous factors do affect the service quality of the token. An example is the collectible game CryptoKitties, which accounted at its peak in December 2017 for 11% of Ethereum's traffic. CryptoKitties'

sharply increased popularity caused bottlenecks for other smart contracts, which rose the transaction fees for blockchain users (Lehnberg, 2018, p. 25). Hence, high network utilization may affect all examined metrics.

Even though this paper applied a time series analysis to avoid spurious regressions, there may be an estimation issue. The time-series data of each token has been examined on whether there exists a cointegrating relationship from the independent variables to the dependent variable. In case there are further cointegrating relationships among the variables, the estimations of the ARDL/ECM may be misspecified. This issue could be addressed with the Vector Error Correction Model (VECM). However, as this would require the time series data to be stationary at the first difference, this would lead to the exclusion of the variables which are stationary at the level form, namely the exchange metrics. Further analysis may consider the utilization of the VECM.

Outlook. As the examination of the market price of crypto assets and token characteristics is an arising topic, there are various opportunities to continue for future research. With respect to the further examination of on-chain metrics, it may be promising to research the behavior of tokens that are predominately exchanged on central platforms. As the token allocation is traceable with the transaction data of the blockchain at every point in time, one could track how the relation of all tokens that are maintained within exchange platforms changes. It is assumed that the price of the assets that are mainly held by exchange addresses behave more sensitive to price shocks than tokens that are predominately allocated to private addresses. Equipped with a token distribution metric, it would be interesting to see if the findings of this paper would differ between those two token groups.

Furthermore, the relationship among exchange addresses might bear more insights. As one can track transactions to and from exchange platforms, it may be insightful to investigate how the price of a token behaves when the exchange of interest is receiving that particular token. A further idea is to analyze pending transactions that are not yet processed by miners, especially when exchange platforms are addressed. A drastic increase in the number of

transactions, respectively, the number of tokens moved to exchanges, could be interpreted as a signal that the market price will decline in the near term.

With respect to a further regression analysis on the market price of crypto assets, the consideration of structural breaks might be fruitful. The analysis for different periods would allow to understand how the impact of different metrics changes over time, for example, pre- and post ICO bubble burst. This facilitates seeing whether pre-tested metrics become relevant at a later stage. As the crypto asset market matures, it becomes more interconnected to other asset classes. Hence, a later examination of the relationship between macro factors or asset classes on asset prices may be prosperous.

Alexander Stober *studied Business Informatics (Bachelor) and Finance (Master) at the Frankfurt School of Finance and Management. You can contact him via mail (alexanderstober@gmail.com).*

Prof. Dr. Philipp Sandner *is head of the Frankfurt School Blockchain Center. You can contact him via mail (email@philipp-sandner.de), via LinkedIn (https://www.linkedin.com/in/philippsandner/) or follow him on Twitter (https://twitter.com/philippsandner).*

Bibliography

Adhami, S., Giudici, G., & Martinazzi, S. (2018). Why do Businesses go Crypto? An Empirical Analysis of Initial Coin Offerings. *Journal of Economics and Business, Vol. 100, No. 6*, pp. 64-75. doi:10.1016/j.jeconbus.2018.04.001.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, Vol. 19, No. 6*, pp. 215-222. doi:10.1109/TAC.1974.1100705.

Alabi, K. (2017). Digital Blockchain Networks Appear to be Following Metcalfe's Law. *Electronic Commerce Research and Applications, Vol. 24*, pp. 23-39. doi:10.1016/j.elerap.2017.06.003.

Ante, L., Sandner, P., & Fiedler, I. (2018). Blockchain-Based ICOs: Pure Hype or the Dawn of a New Era of Startup Financing? *Journal of Risk and Financial Management, Vol. 11, No. 4*, pp. 1-19. doi:10.3390/jrfm11040080.

Baur, D., & McDermott, T. (2010). Is Gold a Safe Haven? International Evidence. *SSRN Electronic Journal*, p. 35. doi:10.1016/j.jbankfin.2009.12.008.

Bheemaiah, K., & Collomb, A. (2018). *Cryptoasset Valuation: Identifying the Variables of Analysis*. Retrieved July 12, 2019, from Louis Bachelier: www.louisbachelier.org/wp-content/uploads/2018/10/cryptovaluationreport-v20181016-fv.pdf.

Blockchain. (2018). *The State of Stablecoins*. Retrieved July 13, 2019, from Blockchain: www.blockchain.com/ru/static/pdf/StablecoinsReportFinal.pdf.

Buterin, V. (2013). *A Next Generation Smart Contract & Decentralized Application Platform*. Ethereum White Paper. Retrieved July 10, 2019, from http://blockchainlab.com/pdf/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf.

Ciaian, P., Rajcaniova, M., & Kancs, d. (2016). The Economies of BitCoin Price Formation. *Applied Economics, Vol. 48, No. 19*, pp. 1799-1815. doi:10.1080/00036846.2015.1109038.

Ciaian, P., Rajcaniova, M., & Kancs, d. (2018). Virtual Relationships: Short- and Long-Run Evidence from Bitcoin and Altcoin Markets. *Journal of International Financial Markets, Institutions & Money, Vol. 52*, pp. 173-195. doi:10.1016/j.intfin.2017.11.001.

CoinMarketCap. (2019). *Top 100 Tokens by Market Capitalization*. Retrieved July 15, 2019, from https://coinmarketcap.com/tokens/.

eth.events. (2019). *Query Ethereum with SQL*. Retrieved July 15, 2019, from eth.events: https://eth.events/.

Etherscan. (2020). *ERC-20 Token Tracker*. Retrieved August 29, 2019, from Etherscan: https://etherscan.io/tokens.

Granger. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometics, Vol. 16, No. 1*, pp. 121-130. doi:10.1016/0304-4076(81)90079-8.

Granger, C., & Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econopmetrics, Vol. 2, No. 2*, pp. 111-120. doi:10.1016/0304-4076(74)90034-7.

Greene, W. H. (2007). *Econometric Analysis, 6th Edition*. New Jersey, USA: Pearson Prentice Hall.

Kripfganz, S., & Schneider, D. C. (2018). *ARDL: Estimating Autoregressive Distributed Lag and Equilibrium Correction Models*. Retrieved July 12, 2019, from Stata: www.stata.com/meeting/uk18/slides/uk18_Kripfganz.pdf.

Lehnberg, D. (2018). Valuing the Top 50 Ethereum ERC-20 Tokens Using Metcalfe's Law. *Unpublished bachelor's thesis, Jönköping University, Jönköping, Sweden*, pp. 1-44.

Li, X., & Chong, A. W. (2017). The Technology and Economic Determinants of Cryptocurrency Exchange Rates: The Case of Bitcoin. *Decision Support Systems, Vol. 95*, pp. 49-60. doi:10.1016/j.dss.2016.12.001.

Metcalfe, R. (2013). Metcalfe's Law after 40 Years of Ethernet. *Computer, Vol. 46, No. 12*, pp. 26-31. doi:10.1109/MC.2013.374.

Nkoro, E., & Uko, A. K. (2016). Autoregressive Distributed Lag (ARDL) Cointegration Technique: Application and Interpretation. *Journal of Statistical and Econometric Methods, Vol. 5, No. 4*, pp. 63-91.

Pesaran, M. H., & Shin, Y. (1999). An Autoregressive Distributed Lag Modeling Approach to Co-Integration Analysis. In S. Strom, *Econometrics and Economic Theory in the 20th century: The Ragnar Frish Centennial Symposium* (pp. 371-413). Cambridge, USA: Cambridge University Press. doi:10.1017/CCOL0521633230.011.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bound Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics, Vol. 16, No. 3*, pp. 289-326. doi:10.1002/jae.616.

Polasik, M., Piotrowaska, A. I., Wisniewski, T. P., Kotkowski, R., & Lightfoot, G. (2015). Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry. *International Journal of Electronic Commerce, Vol. 20, No. 1*, pp. 9-49. doi:10.1080/10864415.2016.1061413.

Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston, Massachusetts, USA: Harvard Business School Press.

Shrestha, M. B., & Bhatta, G. R. (2018). Selecting Appropriate Methodological Framework for Time Series Data Analysis. *The Journal of Finance and Data Science, Vol. 4*, pp. 71-89. doi:10.1016/j.jfds.2017.11.001.

Somin, S., Gordon, G., & Altshuler, Y. (2018). Network Analysis of ERC20 Tokens Trading on Ethereum Blockchain. (pp. 1-12). Cambridge, Mass.: Springer Proceedings in Complexity. doi:10.1007/978-3-319-96661-8_45.

Swann, P. (2002). The Functional Form of Network Effects. *Information Economics and Policy, Vol. 14, No. 3*, pp. 417-429. doi:10.1016/S0167-6245(02)00051-3.

Tokenbase. (2019, July). *Tokenbase*. Retrieved from International Token Standardization Association (ITSA): www.itsa.global.

Tucker, C. (2008). Identifying Formal and Informal Influence in Technology Adoption with Network Externalities. *Management Science, Vol. 54, No. 12*, pp. 2024-2038. doi:10.1287/mnsc.1080.0897.

Victor, F., & Lüders, B. (2019). Measuring Ethereum-based ERC20 Token Networks. *International Conference on Financial Cryptography and Data Security*, pp. 1-17.

Wei, W. C. (2018). The Impact of Tether Grants on Bitcoin. *Economics Letter, Vol. 171, No. 10*, pp. 19-22. doi:10.1016/j.econlet.2018.07.001.

Wheatley, S., Sornette, D., Huber, T., Reppen, M., & Gantner, R. (2018). Are Bitcoin Bubbles Predictable? Combining a Generalized Metcalfe's Law and the LPPLS Model. *Swiss Finance Institute Research Paper, No. 18-22*, pp. 1-21.

Endnotes

[1] For more information regarding the required and optional functions an ERC-20 token has to meet, see Bheemaiah & Collomb (2018, p. 58 ff) and Victor & Lüders (2019, p. 3).
[2] Since $N \times (N-1)$ is asymptotic to N^2, Metcalfe's Law can be written as $V \propto N^2$.
[3] According to Swann (2002) there is also Sarnoff's Law, examining the aggregate utility of one-way communication and Reed's Law, examining the aggregate utility of group forming networks. There has been criticism about the validity of Metcalfe's Law, which was addressed by Swann (2002).
[4] Note that the peg to the underlying commonly fluctuates around the basis price. This issue is in detail discussed in Wei (2018).
[5] However, Baur & McDermott only examined the effects of stock markets on the gold price, but not vice versa. Investors suggesting a relationship between the stock and the gold market may interpret rising gold prices as a signal to sell their shares and buy gold.
[6] We assume that counting the unique addresses over one month will reflect the user base of a crypto asset better than for shorter time frames. This assumption is based on different stakeholder groups. Token buyers may (i) purchase the token for its original intent or use the asset for the (ii) purpose of speculation. On the one hand, the purpose of the utility might not be applicable yet, considering that tokens are initially sold to fund the platforms on which they can be used. This group would retain its token for a more extended period. On the other hand, investors may acquire the token in the hope of quick returns and never get in touch with the asset after the sale. Hence, aggregating the active addresses over a mediocre period may represent a more accurate picture of the different stakeholders than smaller time frames.
[7] The data collection has been conducted with a list of exchange addresses, sourced from Etherscan (2019).
[8] Strictly speaking, this is also true for the transactions of the analyzed ERC-20 tokens, as there are also trades happening on central exchange platforms, which are not part of the ledger and are therefore not accounted in this method. Abandoning tokens which are used on different blockchains for tokens which are only used on Ethereum has the advantage that the gathered data do not needs a further adaption, respectively transformation.
[9] The AIC is based on information theory and used to determine the optimal lag length (Akaike, 1974). The formula of AIC is: $AIC = -2\ln(L) + 2K$, where L is the maximum likelihood function for the estimated time series, and K is the number of estimated parameters.
[10] The second difference, for example, is the first difference of the first difference.
[11] Despite the relevance of the unit-root test for the time series analysis, the ADF test results are briefly summarized and not displayed to provide clarity.
[12] A value of zero implies a positive, a value of four implies a negative, and a value of two implies no autocorrelation of the residuals.