Informational Efficiency of Cryptocurrency Markets

Mahendrarajah Nimalendran, Praveen Pathak, Mariia Petryk, Liangfei Qiu

Internet Appendix

Appendix A. Asset Market Regulation

Table A.1. Comparative Analysis of the Regulatory Institutions and Norms

Table A.1 summarizes the main regulatory requirements for the companies issuing financial assets publicly traded in the U.S.

	SEC	FinCEN
Status	Independent federal government agency in the U.S.	Bureau within the Department of Treasury in the U.S.
Purpose	Protecting investors, maintaining efficient markets, and facilitating capital formation[1]	Combatting financial crimes and promoting national security
Subject of regulation	Securities	Transactions
Character of regulation	Creates the information disclosure policies and enforces their execution	Monitors the transaction flows and prevents money laundering and financing of terrorism
Sample documents and regulatory acts[2]	Securities Act of 1933 [17 CFR Part 230]; Securities Exchange Act of 1934 [17 CFR Part 240]; including Exchange Act registration and reporting (Reg 12B); Securities ownership (Reg 13D and 13G); Integrated disclosure requirement repository Regulation S-K [17 CFR Part 229] etc.	The Currency and Foreign Transactions Reporting Act of 1970; USA PATRIOT Act of 2001; The Anti-Money Laundering Act of 2020; The Corporate Transparency Act (CTA).
Stages of the project the regulation is applied	Public issue of shares or other tokens representing the stake in business in exchange for capital; In the process of economic activity and interactions with stakeholders	Mostly in the process of economic activity and interactions with other subjects.
Points of collision with cryptocurrencies	When investors (U.S. citizens) invest in the assets	When cryptocurrency transactions are used as a part of illegal activity

[1] More information about SEC goals can be found at https://www.sec.gov/our-goals, FinCEN – at https://www.fincen.gov/what-we-do. (Last accessed: March 8, 2024)
[2] The full list of the SEC regulatory documents can be found at https://www.sec.gov/divisions/corpfin/ecfrlinks, FinCEN – at https://www.fincen.gov/fincens-legal-authorities. (Last accessed: March 8, 2024)

Appendix B. Variance Ratio Methodology

This study follows the Lo and MacKinlay (1988) methodology to test for market efficiency. The test statistics are based on the null hypothesis that the stock prices follow a random walk process. The random walk process is a nonstationary process for price levels, for which conditional mean and variance of returns are both linear functions of time. It is usually referred to as Random Walk 1 or $RW1$. Let price be represented as a random walk as in equation (A1):

$$P_t = c + P_{t-1} + \varepsilon_t, \quad P_t \text{ is Price at } t, \varepsilon_t \sim IID(0, \sigma^2). \tag{A1}$$

Given some initial value, P_0, at time zero, and assuming linearity of conditional mean, $E[P_t|P_0] = P_0 + ct$, where condition $c \neq 0$ is assumed to allow for a drift, and variance, $Var[P_t|P_0] = \sigma^2 t$, the random walk process is described in equation (A2):

$$P_t = P_0 + ct + \sum_{i=0}^{t-1} \varepsilon_{t-i} \,. \tag{A2}$$

Here t is the time index. Note that cryptocurrencies typically do not bear any dividend payments. Thus, price changes are the same as returns, unlike interest on dividend-bearing securities.

The assumption of identically distributed increments can be relaxed with independent but not identically distributed (INID) increments ε_t, usually called Random Walk 2 ($RW2$). In contrast with $RW1$ process, $RW2$ allows for unconditional heteroscedasticity in increments ε_t. Random walk assumption about the independence of ε_t can be replaced by the assumption that increments ε_t are uncorrelated to make it more general, which implies that $cov[\varepsilon_t, \varepsilon_{t-k}] = 0$, for all k, but $cov\left[\varepsilon_t^2, \varepsilon_{t-k}^2\right] \neq 0$ is allowed. This process is usually called Random Walk 3, or $RW3$.

Let, $p_t = \ln(P_t)$, then,

$$p_t - p_{t-1} = c + \varepsilon_t, \quad \varepsilon_t \sim IID(0, \sigma^2). \tag{A3}$$

The continuously compounded returns, r_{it}, defined in equation (A4) is used for the VR estimates.

$$r_{it} = \ln\left(\frac{P_{it}}{P_{i(t-1)}}\right) = \ln(P_{it}) - \ln\left(P_{i(t-1)}\right) = p_t - p_{i(t-1),}, for\ asset, i. \tag{A4}$$

Let $r_{it}(q)$ denote the return with q base periods in-between the two individual observations:

$$r_{it}(q) \; = \ln\left(\frac{P_t}{P_{t-q}}\right) = \ln(P_t) - \ln\left(P_{t-q}\right) = p_t - p_{t-q}. \tag{A5}$$

The VR estimator is computed using the following formula (A6):

$$VR_i(q) = \frac{Var\left(r_{it}(q)\right)}{q\sigma^2}, \tag{A6}$$

where $Var(r_{it}(q))$ is the variance of q-period continuously compounded return, σ^2 is the variance of one-period returns, and q is the number of periods.

The null and alternative hypotheses to test for efficiency are described in equation (A7).

$$H_0: VR_i(q) - 1 = 0 \quad vs \quad H_1: VR_i(q) - 1 \neq 0. \tag{A7}$$

If the null hypothesis is rejected, then either autocorrelations or heteroscedasticity of returns are present in the sample, and the prices are inefficient. It can also be shown the VR statistic is a weighted sum of all the autocorrelations up to lag $q-1$, as described in equation (A8).

$$VR_i(q) = \frac{Var\left(r_{it}(q)\right)}{q\sigma^2} = 1 + 2\sum_{k=1}^{q-1}\left(1 - \frac{k}{q}\right)\rho(k), \tag{A8}$$

where, $\rho(k)$, is the autocorrelation coefficient of the k^{th} order of the process $\{r_{it}\}$. Under H_0 all autocorrelation coefficients $\rho(k)$ are equal to zero, hence the $VR_i(q) = 1$. For further analysis, we will denote $M_i(q) \equiv VR_i(q) - 1$.

To account for heteroscedasticity in ε_t, i.e., as described by *RW2*, when the variance is no longer a linear function of the time variable, and $Var[P_t|P_0] \neq \sigma^2 t$, the heteroscedasticity-consistent estimator of the variance of $\hat{\rho}(j)$ given by equation (A9) is used (Lo and MacKinlay 1988).

$$\hat{\delta}(j) = \frac{\sum_{t=j+1}^{nq}(p_{it} - p_{it-1} - \hat{\mu})^2\left(p_{it-j} - p_{it-j-1} - \hat{\mu}\right)^2}{\left(\sum_{t=1}^{nq}(p_{it} - p_{it-1} - \hat{\mu})^2\right)^2}. \tag{A9}$$

The heteroscedasticity-consistent estimator of $Var(M_i(q))$ is given in equation (A10).

$$Var\left(M_i(q)\right) = \hat{\theta}(q) = \sum_{j=1}^{q-1}\left(\frac{2(q-j)}{q}\right)^2 \hat{\delta}(j). \tag{A10}$$

To test the null hypothesis, we use standardized z test statistic, $\hat{z}_{\alpha_i}$, which is asymptotically standard normal (equation A11).

$$\hat{z}_{\alpha_i} = \frac{\sqrt{nq}M_i(q)}{\sqrt{\hat{\theta}(q)}} \sim_a N(0,1). \tag{A11}$$

VR Estimates Based on Overlapping Data

Lo and MacKinlay (1988) suggest using overlapping data to estimate $q - period$ variances to improve the efficiency of the estimator. The $q - period$ estimator is described by equation (A12).

$$Var(r_{it}(q), T) = \frac{1}{m} \sum_{t=q+1}^{T} \left(p_{it} - p_{it-q} - q\hat{\mu}\right)^2, T = 9, 10, \dots T_S. \tag{A12}$$

Here, $\hat{\mu}$ is the mean of one-period returns, and $m = q(T - q + 1)(1 - \frac{q}{T})$. This gives eight daily returns to estimate $Var\left(r_{it}(q)\right)$, when $q = 1\ and\ T = 9$. When $q = 7$, there are two overlapping 7-day returns. Day T is then advanced by one day to estimate the VR for day ten for the offering i. This provides a time series of estimates for each ICO, IEO, and IPO in our sample, $VR(i, T), T = 9\ to\ T_S$. The number of observations T_S is equal to the number of days until the firm gets delisted or the end of the sample period, whichever is longer. The variance ratio is calculated using equation (A13).

$$VR_{i,q,T} = \frac{Var(r_{it}(q), T)}{qVar(r_{it}(1), T)} \tag{A13}$$

For each day, T, we estimate the simple average (or the volume-weighted) across all surviving offerings i using formula (A14):

$$\overline{VR}_{q,T} = \frac{1}{N_S} \sum_{i=1}^{N_S} w_i VR_{i,q,T}. \tag{A14}$$

Here N_S is the number of surviving firms on the day T, and w_i is the volume weight for firm i, or equal weighting.

VR Parameter q Choice

According to the EMH, the asset market price reflects the change in information regularly and automatically. In case of market sentiment changes, informed investors take immediate action trading the asset. However, the speed of price change delivery may not be immediate for the less informed groups of investors, e.g., the next day in mass media outlets. Therefore, we use the daily return as the base to reflect the market reaction. In addition, the prior research uses the weekly return as the base for variance ratio calculation (Lo and McKinlay (1988), Liu and He (1991)) to avoid inconveniences associated with the non-trading periods in the stock markets. The cryptocurrency markets are open 24/7, which alleviates this concern. In addition, the weekly business cyclicity makes the weekly interval meaningful from a portfolio performance-measurement perspective.

Appendix C. Supplementary Figures and Tables for Analysis

Table C.1. Descriptive Statistics of Key Analysis Variables

Table C.1 provides the descriptive statistics summary of the key variables for all asset samples in the analysis. Columns 1-4 and 5-8 present the number of observations, mean, median, and standard deviation of the two matched samples respectively.

Variables	Sample 1				Sample 2			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N	Mean	Median	SD	N	Mean	Median	SD
	IPO				**ICO**			
VR	9,975	0.908	0.878	0.307	5,611	0.89	0.90	0.28
DailySpread	9,975	0.022	0.005	0.036	5,611	0.03	0.01	0.05
DailyTurnover	9,975	17.579	4.446	185.072	5,611	0.20	0.03	0.67
DaysTraded	9,975	155.92	145.00	96.37	5,611	202.03	211.00	99.64
	IPO				**IEO-L**			
VR	8,861	0.874	0.859	0.35	7,990	0.964	0.943	0.31
DailySpread	8,861	0.025	0.005	0.05	7,990	0.028	0.002	0.05
DailyTurnover	8,861	38.014	4.119	388.95	7,990	1.238	0.175	4.47
DaysTraded	8,861	137.728	125	87.32	7,990	191.362	192	101.10
	ICO				**IEO-NL**			
VR	7,018	0.905	0.933	0.279	5,476	0.97	0.96	0.338
DailySpread	7,018	0.031	0.005	0.05	5,476	0.03	0.003	0.049
DailyTurnover	7,018	0.176	0.03	0.61	5,476	1.902	0.027	6.223
DaysTraded	7,018	199	207	100	5,476	204	215	100
	IEO-L				**IEO-NL**			
VR	7,990	0.964	0.943	0.31	3,394	0.891	0.774	0.37
DailySpread	7,990	0.028	0.002	0.05	3,394	0.032	0.004	0.06
DailyTurnover	7,990	1.238	0.175	4.47	3,394	0.031	0.003	0.11
DaysTraded	7,990	191	192	101	3,394	200	203	98
	IDO				**IEO**			
VR	1,107	0.876	0.727	0.28	4,290	0.919	0.887	0.366
DailySpread	1,107	0.031	0	0.06	4,290	0.027	0.002	0.043
DailyTurnover	1,107	0.008	0.003	0.04	4,290	2.178	0.689	5.920
DaysTraded	1,107	214.156	227	96.33	4,290	187.112	187.000	102.740

Table C.2. Model of Market Efficiency with Maturity Interaction Term

Table C.2 shows the autocorrelation-corrected robust estimates of the model with interaction term $D * DaysTraded$ explaining the relationship of trading platform supervision between the cryptocurrency and its market inefficiency. Columns 1-5 and 6-10 present the model estimation results for H1-H5 in the short term and long term respectively. The ***, **, and * denote statistical significance at the 1%, 5%, and 10% level correspondingly. Robust t-statistics clustered at the asset level are presented in parentheses.

Panel A: 100 days					
Variable	(1) H1: IPO and ICO	(2) H2: IPO and IEO-L	(3) H3: IEO-NL and ICO	(4) H4: IEO-L and IEO-NL	(5) H5: IDO and IEO
D_{IPO}	0.024				
	(0.79)				
D_{IPO}* DaysTraded	-0.001***				
	(-3.497)				
D_{IPO}		0.032			
		(0.623)			
D_{IPO}* DaysTraded		-0.001			
		(-0.873)			
D_{IEO-NL}			0.064		
			(1.248)		
D_{IEO-NL}* DaysTraded			-0.001*		
			(-1.743)		
D_{IEO-L}				-0.136	
				(-1.49)	
D_{IEO-L}* DaysTraded				-0.0003	
				(-0.205)	
D_{IDO}					-0.058
					(-0.938)
D_{IDO}* DaysTraded					-0.002***
					(-2.862)
DailySpread	0.105	-0.058*	-0.072	0.064	0.269
	(1.614)	(-1.654)	(-0.952)	(1.348)	(1.236)
DailyTurnover	-0.00008***	-0.00003***	0.003	0.011***	0.008**
	(-4.576)	(-4.355)	(1.435)	(2.827)	(2.383)
DaysTraded	-0.003***	-0.003***	-0.003***	-0.002***	-0.002**
	(-4.722)	(-5.768)	(-6.444)	(-1.999)	(-2.548)
Constant	0.560***	0.497***	0.600***	0.596***	0.477***
	(12.865)	(16.749)	(14.973)	(7.189)	(11.975)
Observations	4,631	5,414	2,606	2,557	1,272
R-squared	0.15	0.08	0.157	0.125	0.132
Clustered T-stat	YES	YES	YES	YES	YES

Table C.2. Model of Market Efficiency with Maturity Interaction Term (contd.)

Panel B: 365 days					
Variable	(6) H1: IPO and ICO	(7) H2: IPO and IEO-L	(8) H3: IEO-NL and ICO	(9) H4: IEO-L and IEO-NL	(10) H5: IDO and IEO
D_{IPO}	-0.047*				
	(-1.882)				
D_{IPO}* DaysTraded	0.00004				
	(0.407)				
D_{IPO}		0.056***			
		(2.591)			
D_{IPO}* DaysTraded		-0.0003**			
		(-2.561)			
D_{IEO-NL}			-0.013		
			(-0.919)		
D_{IEO-NL}* DaysTraded			0.0003***		
			(3.955)		
D_{IEO-L}				-0.177***	
				(-7.935)	
D_{IEO-L}* DaysTraded				0.0003***	
				(3.927)	
D_{IDO}					-0.178***
					(-4.468)
D_{IDO}* DaysTraded					0.001***
					(4.94)
DailySpread	0.073**	-0.038	0.048	0.086**	-0.073
	(2.114)	(-1.003)	(1.35)	(2.634)	(-0.774)
DailyTurnover	-0.00002*	-0.00003***	0.003*	0.008***	0.006***
	(-1.67)	(-5.127)	(1.822)	(3.9)	(2.647)
DaysTraded	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(-8.867)	(-6.503)	(-7.834)	(-7.694)	(-10.801)
Constant	0.414***	0.343***	0.409***	0.498***	0.413***
	(18.954)	(16.37)	(14.186)	(18.736)	(25.464)
Observations	15,586	16,851	12,494	11,384	5,397
R-squared	0.166	0.091	0.188	0.164	0.113
Clustered T-stat	YES	YES	YES	YES	YES

Table C.3. Regulation Moderation Effect: ICO and IEO-L

Table C.3 reports the model (5) results for the two matched samples of ICO and IEO-L. Columns 1 and 3 present the OLS heteroskedasticity-consistent estimates for the 100 and 365 days respectively. Columns 2 and 4 present the OLS autocorrelation- and heteroskedasticity-consistent estimates for the 100 and 365 days respectively. The ***, **, and * denote statistical significance at the 1%, 5%, and 10% level correspondingly. Robust and autocorrelation-corrected t-statistics are in parentheses.

Variable	(1) OLS w heteroskedasticity-consistent errors	(2) OLS w autocorrelation-consistent errors	(3) OLS w heteroskedasticity-consistent errors	(4) OLS w autocorrelation-consistent errors
	100 days		365 days	
D_{IEO-L}	-0.068***	-0.068***	-0.003	-0.003
	(-6.221)	(-3.039)	(-0.668)	(-0.336)
DailySpread	0.269***	0.269***	0.113***	0.113***
	(2.749)	(2.497)	(2.769)	(2.225)
DailyTurnover	0.010***	0.010***	0.007***	0.007***
	(3.551)	(2.67)	(4.154)	(3.371)
DaysTraded	-0.003***	-0.003***	-0.001***	-0.001***
	(-11.924)	(-7.723)	(-32.085)	(-7.557)
Constant	0.567***	0.567***	0.378***	0.378***
	(32.098)	(21.087)	(66.01)	(16.098)
Observations	2,604	2,604	11,782	11,782
R-squared	0.112	0.112	0.141	0.141
Clustered SE	YES	YES	YES	YES

Table C.4. Hypothesis Testing Results

Table C.4 presents the summary of hypothesis testing results based on regression analysis.

Hypothesis Testing Result	Hypothesis Text	Result Description
	Short-term (100 days)	
H1+	***Hypothesis 1****: Regulation compliance leads to higher market efficiency, and cryptocurrencies that are not regulated (ICOs) are less efficient than regulated stocks (IPOs).*	Coefficient is negative and significant. IPOs are significantly less inefficient than comparable ICOs.
H2A	**Hypothesis 2A**: *FinCEN and "hard" regulations provide similar levels of transparency and investor protection. Hence, there is no difference in informational efficiency between the two asset classes.*	Coefficient is not significant. The difference in market efficiency between IPOs and IEOs is not significant.
H3A	**Hypothesis 3A:** *Crypto exchange mediation does not affect price efficiency. Hence, ICOs and IEOs are equally informationally efficient.*	Coefficient is not significant. The difference in market efficiency between IEOs underwritten by an exchange without license and not-underwritten ICOs is not significant.
H4+	**Hypothesis 4:** *Crypto assets underwritten by the licensed centralized exchange (IEO-L) are more efficient than crypto assets that are mediated by a third-party platform without a license (IEO-NL).*	Coefficient is negative and significant. IEOs underwritten by an exchange with an MSB license are significantly less inefficient than comparable IEOs supervised by the non-licensed platform.
H5+	**Hypothesis 5:** *The assets that are traded on more liquid markets (IDO) are more efficient than the assets traded on the centralized platforms (IEO).*	Coefficient is negative and significant. IDOs are significantly more efficient than comparable IEOs.
	Long-term (365 days)	
H1+	***Hypothesis 1****: Regulation compliance leads to higher market efficiency, and cryptocurrencies that are not regulated (ICOs) are less efficient than regulated stocks (IPOs).*	Coefficient is negative and significant. IPOs are significantly less inefficient than comparable ICOs.
H2-	***Hypothesis 2****: Crypto assets that are compliant FinCEN are more inefficient than assets compliant with SEC "hard" regulation.*	Coefficient is positive and significant. In the long term, IPOs are significantly more inefficient than comparable FinCEN-licensed IEOs.
H3-	**Hypothesis 3:** *Crypto assets that are mediated by the third-party exchange (IEOs) are more efficient than crypto assets not underwritten by a centralized platform (ICOs).*	Coefficient is positive and significant. IEOs supervised by a platform without an MSB license are significantly more inefficient than comparable ICOs.
H4+	**Hypothesis 4:** *Crypto assets underwritten by the licensed centralized exchange (IEO-L) are more efficient than crypto assets that are mediated by a third-party platform without a license (IEO-NL).*	Coefficient is negative and significant. IEOs underwritten by an exchange with an MSB license are significantly less inefficient than comparable IEOs supervised by the non-licensed platform.
H5A	***Hypothesis 5A****: Liquidity mediation does not affect price efficiency; Hence, IDOs and IEOs are equally informationally efficient.*	Coefficient is not significant. Difference between IDOs and IEOs is not significant.

Table C.5. Pooled Regression Dummy Variables Definitions

Table C.5 presents the dummy variables definitions in the pooled regression.

Dummy	Description
D_{IPO}	IPOs that are in the matched sample with ICO
D_{IEO-L}	IEOs that are in the matched sample with IEO-NL
D_{IEO-NL}	IEOs that are in the matched sample with ICO
D_{IDO}	IDOs on Binance Smart Chain that are in the matched sample with IEO on Binance Smart Chain
D_{IEO}	IEOs on Binance Smart Chain that are in the matched sample with IDO on Binance Smart Chain

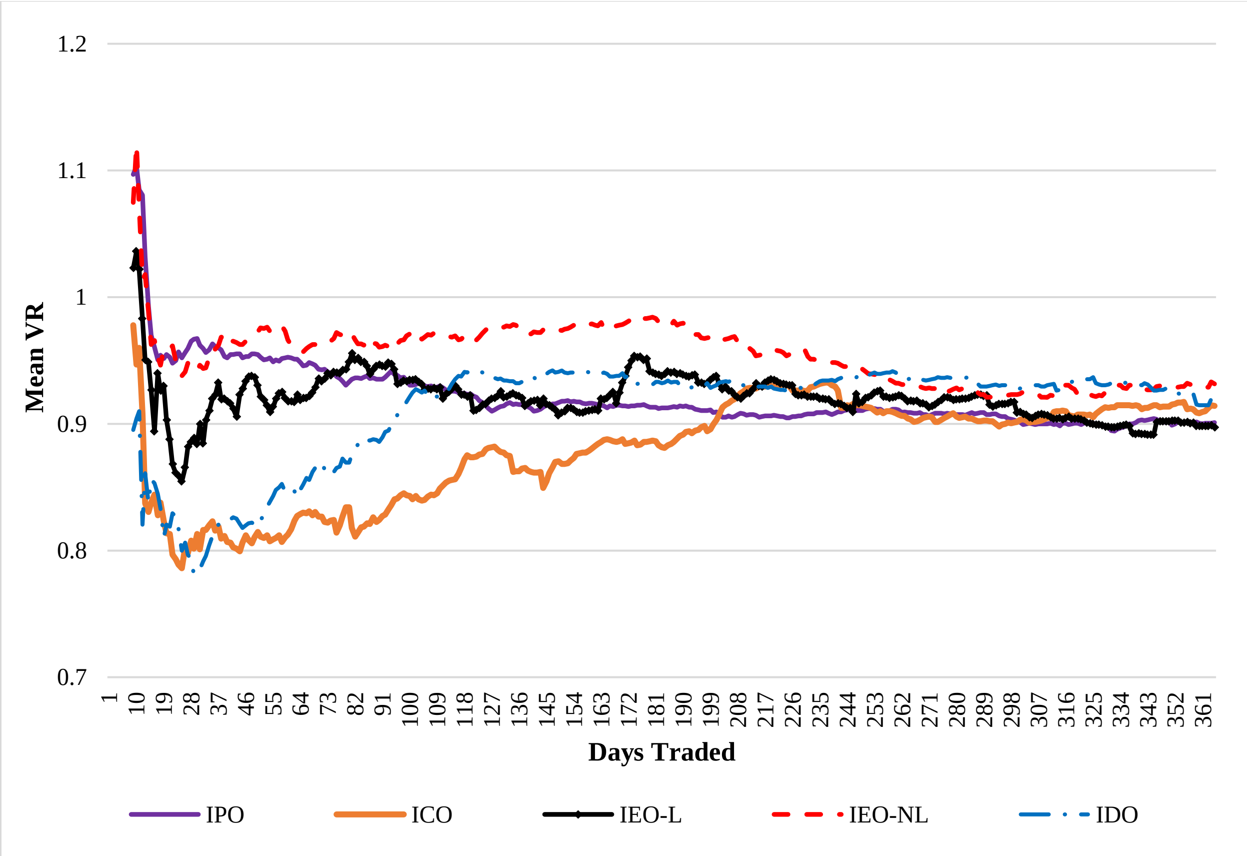
Table C.6. Correlation of Variables

Table C.6 provides the Pearson correlations of the key variables. The ***, **, and * denote statistical significance at the 1%, 5%, and 10% level correspondingly.

	Inefficiency	DailySpread	DailyTurnover	DaysTraded
IPO and ICO sample				
Inefficiency	1	0.048***	-0.004	-0.397***
DailySpread	0.048***	1	0.025***	-0.052***
DailyTurnover	-0.004	0.025***	1	-0.045***
DaysTraded	-0.397***	-0.052***	-0.045***	1
IPO and IEO-L sample				
	Inefficiency	DailySpread	DailyTurnover	DaysTraded
Inefficiency	1	0.006	-0.016**	-0.295***
DailySpread	0.006	1	0.098***	-0.071***
DailyTurnover	-0.016**	0.098***	1	-0.060***
DaysTraded	-0.295***	-0.071***	-0.060***	1
IEO-NL and ICO sample				
	Inefficiency	DailySpread	DailyTurnover	DaysTraded
Inefficiency	1	0.059***	0.108***	-0.410***
DailySpread	0.059***	1	0.072***	-0.108***
DailyTurnover	0.108***	0.072***	1	-0.089***
DaysTraded	-0.410***	-0.108***	-0.089***	1
IEO-L and IEO-NL sample				
	Inefficiency	DailySpread	DailyTurnover	DaysTraded
Inefficiency	1	0.070***	0.138***	-0.298***
DailySpread	0.070***	1	0.037***	-0.110***
DailyTurnover	0.138***	0.037***	1	-0.091***
DaysTraded	-0.298***	-0.110***	-0.091***	1
IDO and IEO sample				
	Inefficiency	DailySpread	DailyTurnover	DaysTraded
Inefficiency	1	0.021	0.167***	-0.266***
DailySpread	0.021	1	0.075***	-0.102***
DailyTurnover	0.167***	0.075***	1	-0.110***
DaysTraded	-0.266***	-0.102***	-0.110***	1
Pooled sample				
	Inefficiency	DailySpread	DailyTurnover	DaysTraded
Inefficiency	1	0.049***	-0.005	-0.340***
DailySpread	0.049***	1	0.014***	-0.076***
DailyTurnover	-0.005	0.014***	1	-0.036***
DaysTraded	-0.340***	-0.076***	-0.036***	1

Figure C.1. VRs Across All Non-Matched Samples

Figure C.1 shows the dynamics of the cross-sectional means of the variance ratio for all asset samples in the analysis. The VR was calculated for each coin or stock, and then the VRs were averaged cross-sectionally for every day (based on $k=1$ day and $q=7$ days). We estimate weekly returns on the 8th day and VRs on the 9th day only. The first few observations of VR have large standard errors due to a limited number of data points. All observations before matching were considered in each sample. To alleviate the impact of outliers, the lowest 2.5% and highest 2.5% of data are winsorized.



Appendix D. Investor Attention and Market Efficiency

In addition to the regulatory and liquidity mechanisms analyzed in the main paper, investor information-seeking behavior may contribute to the market efficiency of the cryptocurrency markets. Halim et al. (2019) used an experimental framework to examine the role of social media in information acquisition. According to his findings, social communication facilitates a larger fraction of the information available in the market and is reflected in asset prices. Other studies have documented that social communication among traders improves market efficiency (Colla and Mele (2010), Han and Yang (2013), Ozsoylev and Walden (2011)).

The lack of centralized information disclosure procedures has led to social media platforms being the primary medium for information dispensed by cryptocurrency projects and crypto experts. Social media platforms that facilitate community interactions, such as Reddit, have become a primary source of information for crypto investors. Logically, we would expect that easy access to social media outlets would increase the number of informed investors and, thus, the information efficiency of cryptocurrency.

On the other hand, it can be argued that cryptocurrencies with a higher level of social media activities might be less efficient because of the higher population of uninformed traders among investors. Shleifer and Summers (1990) argued that attention-grabbing traders are likely uninformed or noisy traders. Therefore, increased coverage of an instrument can lead to more *inefficient* prices. Consequently, the impact of social media on market efficiency is not straightforward, and we test it using the following hypotheses.

> ***Hypothesis D1:*** *Market inefficiency decreases as the number of active users in cryptocurrency communities increases.*
>
> ***Hypothesis D1A****: Market inefficiency does not change as the number of active users in cryptocurrency communities increases.*

Our study uses quantitative social media indicators as proxies for attention to identify their roles in market efficiency. An efficient market requires at least some investors to pay close *attention* to acquiring and processing information and trading to eliminate arbitrage opportunities. Therefore, we used the number of active users (*RedditActiveUsers*) as a proxy for informed traders. Our analysis shows that a higher number of active users is positively associated with the inefficiency (Table D1, column 2), thus, rejecting HD1. One possible explanation for our results is that the number of active social media users is a weak instrument for the informed investors, therefore, in the short term (column 1) we do not find any significance, and in the long term (column 2) the theoretical prediction does not hold.

Table D.1. Pooled Regression with Social Media Variables

DV winsorized. The first 14 days were ommitted due to extreme values. The ***, **, and * denote statistical significance at the 1%, 5%, and 10% level correspondingly. Robust and autocorrelation-corrected t-statistics are indicated in parentheses.

Variable	(1) DV: I =|1- VR|	(2) DV: I =|1- VR|
	100 days	365 days
D_{IEO-L}	-0.226***	-0.213***
	(-8.309)	(-8.276)
D_{IEO-NL}	0.198***	-0.049
	(7.251)	(-1.294)
D_{IDO}	-0.286***	0.234***
	(-9.647)	(7.017)
D_{IEO}	0.501***	0.580***
	(9.727)	(28.381)
DailySpread	-0.101*	-0.030
	(-2.299)	(-1.035)
DailyTurnover	0.0004	0.003*
	(0.268)	(1.895)
DaysTraded	-0.003***	-0.001***
	(-7.506)	(-6.651)
lnRedditActiveUsers	0.002	0.022***
	(0.284)	(2.881)
lnRedditUserComments	-0.600*	-0.032***
	(-1.862)	(-5.896)
Constant	0.577***	0.377***
	(16.070)	(11.340)
Currency FE	YES	YES
Observations	4,824	23,266
R-squared	0.629	0.608
Asset Clustered SE	YES	YES

To test the alternative instrument for the level of informativeness of the investors, we use the number of Reddit comments (*RedditUserComments*) as a proxy for the attention-grabbing capability of cryptocurrencies. Barber and Odean (2008) argue that retail investors are net buyers of *attention-grabbing* stocks since the average investor has limited resources to evaluate many assets when making purchase decisions. Because attention is a scarce cognitive resource (Kahneman (1973)) and purchasing decisions can be resource-intensive, investors are likely to buy assets with attention-grabbing media coverage. We expect that cryptocurrencies with a higher level of comment activities will be more efficient than those with a lower activity level. Consequently, our hypotheses are as follows.

> ***Hypothesis D2:*** *Market inefficiency decreases as the number of social media comments in cryptocurrency communities increases.*
>
> ***Hypothesis D2A****: Market inefficiency increases as the number of social media comments in cryptocurrency communities increases.*

We find that with increased social media comments, the market inefficiency of cryptocurrencies decreases, which confirms our HD2. Our finding is consistent with the attention-grabbing mechanism prediction.

References:

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 785–818.

Colla, P., & Mele, A. (2010). Information Linkages and Correlated Trading. *The Review of Financial Studies*, 23(1), 203–246.

Halim, E., Riyanto, Y. E., & Roy, N. (2019). Costly Information Acquisition, Social Networks, and Asset Prices: Experimental Evidence. *The Journal of Finance*, 74(4), 1975–2010.

Han, B., & Yang, L. (2013). Social Networks, Information Acquisition, and Asset Prices. *Management Science*, 59(6), 1444–1457.

Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.

Ozsoylev, H. N., & Walden, J. (2011). Asset Pricing in Large Information Networks. *Journal of Economic Theory*, 146(6), 2252–2280.

Shleifer, A., & Summers, L. H. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, 4(2), 19–33.

Appendix E. Market Microstructure Biases

Returns based on transaction prices can lead to a bid-ask bounce, causing a negative serial correlation in returns at lag-one and inflated variance estimates (Roll (1984)), thus biasing the variance ratios we estimate. We can eliminate these biases by using mid-quotes (or bid-bid prices) to calculate returns. However, our daily data from *coinmarketcap.com* is a volume-weighted average of the prices from several exchanges. Hence, we expect the prices to be unaffected by the bid-ask bounce, as the likelihood of the prices always being at the bid or ask is very small or attenuated. To confirm that transaction data from crypto exchanges approximate the daily aggregated data from coinmarketcap.com, we estimate the model given in equation (E1) for a subset of Bitcoin quotes and transaction price data.

$$R_{bid2bid} = \beta_0 + \beta_1 R_{close2close} + \varepsilon \qquad \text{(E1)}$$

In equation (E1), $R_{bid2bid}$ is the return based on bid-bid prices obtained from *coinbase.com* and $R_{close2close}$ is the return calculated from daily *coinmarketcap.com* prices. Under the hypothesis that the daily aggregate data provide unbiased estimates, we expect $\beta_0 = 0$, $\beta_1 = 1$. We collected a sample of daily observations of BTC-USD prices from *coinmarketcap.com* from November 2, 2018, to December 11, 2018, and end-of-day bid quotes from *coinbase.com* for the same period.

We find that the estimated value of $\beta_0 = 0.001$ $(t = 0.098)$ is not statistically different from zero, and $\beta_1 = 1.114$ $(t = 10.1)$ is not different from one $(t = 1.10)$ for testing the null $\beta_1 = 1$. The evidence indicates that VRs based on $R_{close2close}$ will not be adversely affected by microstructural biases.

References:

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance*, 39(4), 1127–1139.

Appendix F. Historical Analysis of the Market Efficiency

Figure F.1. Intertemporal Dynamics of Cross-Sectional Mean VRs for ICOs and IPOs

Figure F.1 shows the mean variance ratios of ICOs and IPOs for each day (based on *k=1* day and *q=7* days) up to 1,999 days from the first day of trading. The VR was calculated for each coin or stock, and then the VRs were averaged cross-sectionally for every day. We estimate weekly returns on the 8th day and VRs on the 9th day only. The first few observations of VR have large standard errors due to a limited number of data points. The average VR is based on estimates that are winsorized at 1.5%. The red horizontal line is drawn at VR=1 and signifies the market efficiency level. Our analysis shows that cryptocurrency markets are highly inefficient over long periods, and stock markets are efficient within a short period.

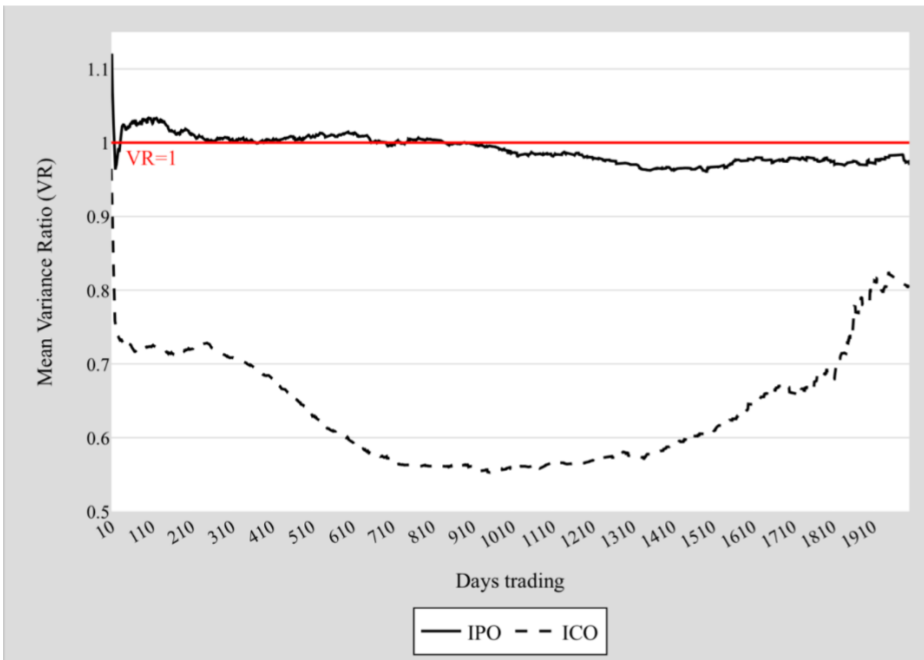


Figure F.2. Intertemporal Dynamics of Cross-Sectional Mean VRs for ICOs and IEOs for the First 210 days of Trading

Figure F.2 shows the mean variance ratios of ICOs and IEOs for each day (based on $k=1$ day and $q=7$ days) up to 210 days from the first day of trading. The VR was calculated for each coin or stock, and then the VRs were averaged cross-sectionally for every day. We estimate weekly returns starting on the 8th day and VRs on the 9th day. The first few observations of VR have large standard errors due to a limited number of observations. The average VR is based on estimates that are winsorized at 1.5%. The red horizontal line is drawn at VR=1 and signifies the market efficiency level.