User participation in cryptocurrency derivative markets

Daisuke Kawai
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Bryan Routledge
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Kyle Soska
Ramiel Capital, New York, New York, USA

Ariel Zetlin-Jones
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Nicolas Christin
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Abstract

As cryptocurrencies have been appreciating against fiat currencies, global markets for cryptocurrency investment have started to emerge, including, most prominently, derivative exchanges. Different from traditional derivative markets, cryptocurrency derivative products are directly marketed to consumers, rather than through brokerage firms or institutional investors. Cryptocurrency derivative exchange platforms include many game-like features (e.g., leaderboards, chatrooms, loot boxes), and have successfully attracted large numbers of investors. This paper attempts to discover the primary factors driving users to flock to these platforms. To answer this question, we have collected approximately a year worth of user data from one of the leading cryptocurrency derivative exchanges between 2020 and 2021. During that period, more than 7.5 million new user accounts were created on that platform. We build a regression analysis, accounting for the idiosyncrasies of the data at hand – notably, its non-stationarity and high correlation – and discover that prices of two major cryptocurrencies, Bitcoin and Ethereum, impact user registrations both in the short and long run. On the other hand, the influence of a less prominent coin, Ripple, and of a "meme" coin with a large social media presence, Dogecoin, is much more subtle. In particular, our regression model reveals the influence of Ripple prices vanishes when we include the SEC litigation against Ripple Labs, Inc. as an explanatory factor. Our regression analysis also suggests that the Chinese government statement regarding tightening cryptocurrency mining and trading regulations adversely impacted user registrations. These results indicate the strong influence of regulatory authorities on cryptocurrency investor behavior. We find cryptocurrency volatility impacts user registrations differently depending on the currency considered: volatility episodes in major cryptocurrencies immediately affect user registrations, whereas volatility of less prominent coins shows a delayed influence.

2012 ACM Subject Classification General and reference → Measurement; Applied computing → Digital cash

Keywords and phrases Cryptocurrency, Online Markets, Derivatives, Trading, Regression Analysis

Digital Object Identifier 10.4230/LIPIcs.AFT.2023.8

Funding This research was partially supported by Ripple's University Blockchain Research Initiative (UBRI) at Carnegie Mellon and by the Carnegie Mellon CyLab Secure Blockchain Initiative. Some of the authors hold non-negligible cryptocurrency positions, but none on the platform under study.
Daisuke Kawai: DK is supported by the Japanese Government Long-Term Overseas Fellowship Program.

1 Introduction

Cryptocurrencies have had a growing impact on global finance. Shortly after the emergence of Bitcoin [33], use cases were primarily as a payment instrument for online fringe activities such as gambling, or the purchase of illegal goods [11, 31]. However, spot prices (i.e., the exchange rate to fiat currencies) rapidly skyrocketed — Bitcoin went from being worth nothing in 2009 to exceeding $60,000 in 2021 – so that cryptocurrencies became an important type of (speculative) financial asset [16].

Consequently, trading infrastructure rapidly expanded from spot exchanges, where people exchange cryptocurrencies for fiat currencies [32], to cryptocurrency derivative platforms [44]. Today, approximately 50–100 billion US dollars are traded every day on these off-chain derivative exchanges.[1] This number far exceeds that of cryptocurrency spot markets, and can be compared to the roughly 200 billion USD traded on the NASDAQ on a given day at the time of writing.[2] In short, cryptocurrency derivative markets are critical to understand the impact of cryptocurrencies on global finance.

The rapid increase in trading volume and user participation led financial regulators to pay close attention. The U.S. Securities and Exchange Commission (SEC) Chair famously emphasized the need for stronger regulations for better investor protection and market integrity [49]. At the international level, the Financial Stability Board (FSB) raised its risk evaluation of cryptocurrency and prioritized the risk assessment of cryptocurrency markets for 2022 [16, 17]. Out of these concerns about potential threats, financial authorities took regulatory measures regarding the cryptocurrency industry [5, 21, 50].

These regulatory changes, as well as large price swings, are expected to impact investor behavior. However, little quantitative analysis has been conducted to measure the degree of influence of all of these potential factors. The core contribution of this paper is to examine the degree to which price appreciation, volatility, and regulatory measures influence user decisions to engage in cryptocurrency investments. To do so, we rely on a dataset we obtained about the hourly performance data of more than eight million investors (registered by July 20, 2021, and most of whom are presumed to be invididual investors) in one of the largest cryptocurrency derivatives markets, from which we can derive how many new investors sign up to the exchange. We use that data to investigate how cryptocurrency prices affect the number of investors in the market with a regression model that can address the long-run relationship between the new registration and major cryptocurrency prices.

A prevailing narrative is that short-term speculation motivates cryptocurrency investments [15, 29] – if so, investors should flock to investment platforms as market volatility increases. We look at the effect of four cryptocurrencies ("reserve" cryptocurrencies like Bitcoin, "meme" currencies like Dogecoin, etc.) prices and volatility on investor registrations, and build a regression to tease out factors that appear to matter. Building this regression presents a number of technical challenges we elaborate on, and our analysis ultimately shows a nuanced picture. The number of investors increases over time, with both price rise and volatility acting as a crucial effect on the rate of increase. However, not all currencies are equal: contrary to Bitcoin or Ethereum, whose price hike and high volatilities immediately affect user registration, Ripple and Dogecoin prices have much less impact on user registrations in the short term, and it takes longer time for the impact of their high volatility to materialize. Our regression also shows the significant influence of regulatory measures. Our analysis shows

[1] https://coinalyze.net/futures-data/global-charts/
[2] https://www.nasdaqtrader.com/Trader.aspx?id=DailyMarketSummary

that the SEC litigation against Ripple Labs, Inc. and its executives basically negated any positive effect of Ripple's price rise on user registrations. The same analysis also suggests that the statement by the Chinese government that it was tightening cryptocurrency regulation[3] also adversely affected user registrations.

2 Related work

Bitcoin is a digital asset maintained by cryptographic primitives and distributed ledger technology. All transactions are recorded on a public ledger ("blockchain") and verified by peers engaging in a cryptographic puzzle ("miners"). Originally proposed as a payment method independent of trusted third parties [33], Bitcoin's use cases during its first few years were fraught with controversy: Meiklejohn et al. [31] showed that one of the major outlets for Bitcoin transactions was Silk Road, a marketplace for (mostly) illegal goods [11]. Moore and Christin showed that Bitcoin exchanges, where people trade Bitcoin for national ("fiat") currencies, frequently failed, and sometimes absconded with their users' money [32].

Despite (or maybe thanks to) the negative publicity, Bitcoin price skyrocketed within a few years. Multiple pieces of literature tried to understand why. Kristoufek [27] showed a correlation between Bitcoin price and the volume of related online search queries. In addition, they found that increased interest in Bitcoin inflates its price, which leads to a bubble-like price movement. Ciaian et al. [12] showed that Bitcoin's attractiveness to investors is an important driver, along with other conventional economic determinants. Urquhart [45] showed that an increase in realized Bitcoin price volatility is correlated to a larger number of related online searches one day later.

More generally, researchers proposed theoretical foundations to integrate various price determinants that had been observed empirically [7, 13, 34, 35, 40, 43]. Network effects appear critical: cryptocurrency appeal, and thus price, grows with the number of users, due to the increased security and (indirectly) usability a large user base provides. For instance, Liu and Tsyvinski's recent empirical analysis [28] shows that cryptocurrency prices correlate with the growth in the number of active on-chain addresses.

By analyzing conditional exposure to tail risks in other cryptocurrencies and in conventional financial asset prices, Borri [8] had showed cryptocurrency prices were affected by other cryptocurrencies, but were decoupled from conventional financial assets prices. Iyer [22] argues this may no longer be the case: correlation between cryptocurrency prices and conventional financial asset prices has been growing.

While this growing body of literature looks into correlations between cryptocurrencies and other financial assets, relatively little is known about market participants. Baur et al. [6] analyzed early Bitcoin holder demographics between 2011 and 2013 and showed that the main purpose of holding Bitcoin is for investment. By analyzing the BitMEX platform, Soska et al. [44] showed derivative investors were a mix of hobbyists and professional traders—with the latter often winning against the former. Kawai et al. [26] show that some derivatives investors provide unreliable investment advice on Twitter.

Despite these advances, many critical issues to characterize cryptocurrency investor behavior are yet to be answered. One of the issues is the influence of the price of major cryptocurrencies on potential investors – i.e., people who have not yet opened investment accounts in cryptocurrency markets, but are interested in investing. We argue

[3] `https://www.gov.cn/xinwen/2021-05/21/content_5610192.htm?ivk_sa=1023197a`

this understanding is critical to better constructing a sustainable cryptocurrency investment environment.

3 Dataset

We obtained investor performance records over two years and a half from a large cryptocurrency derivative exchange public API, and use a subset of this data in the present paper. This section first briefly describes *perpetual futures*, the derivatives product predominantly traded on the exchange, before discussing the investor data present in our dataset.

3.1 Cryptocurrency derivative exchanges

While several platforms investigated various types of cryptocurrency contracts, BitMEX is generally credited with pioneering cryptocurrency derivative products, starting in November 2014 [2,3]. Compared to conventional derivatives markets, the most popular contract available is the *perpetual futures* contract, which, contrary to conventional derivative products (e.g., options), has *no expiry date*: Investors can hold their positions as long as their margin size is large enough to avoid liquidation. Soska et al. present a comprehensive study of BitMEX and of the perpetual futures contract [44]. Below we provide a quick summary of this type of contract, which subsequently became highly popular on all derivative exchanges, including the one we study in this paper.

3.1.1 Perpetual futures

Perpetual futures are investments in the future value of underlying cryptocurrencies: a typical case is the value of Bitcoin (BTC) against US dollar (USD) – or a related "stablecoin" (a cryptocurrency pegged to a fiat currency) like Tether (USDT). Investors of perpetual futures can go "long" or "short." An investor expecting a rise in BTC value against USD will go long (i.e., bet on the appreciation of BTC); conversely, investors expecting a decline will go "short." Longs and shorts are evenly matched among investors: every long contract is paired with a corresponding short contract placed by other investors.

Perpetual cryptocurrency future markets typically allow very high leverage, far beyond what their traditional finance counterparts tolerate. For instance, BitMEX [44] allowed up to 100x leverage. The platform we study allowed up to 125x leverage during the period we investigate (September 2020–July 2021). In short, an investor could invest up to 125 BTC worth of USD with only 1 BTC worth of USD as collateral. If the investor goes long (resp. short), and the value of bitcoin appreciates (resp. depreciates) against the US dollar, the investor can reap significant profit. On the other hand, leveraged positions are incredibly risky: for a 125x leveraged position, a swing of (slightly less than)[4] 0.8% compared to the purchase price, in the direction opposed to the bet made, results in liquidation. That is, the investor's position is immediately closed, and the investor loses all their money.

3.1.2 Performance indices

The exchange we study uses two indices to characterize investor performance: *Profit and Loss* (PnL) and *Return on Investment* (RoI). PnL shows the absolute profit (resp. loss) of an investment portfolio. An absolute metric, PnL tends to get large with investors who can take

[4] Due to transaction fees and other early liquidation mechanisms.

larger positions. On the other hand, the RoI, defined as the PnL divided by the investors' margin size (i.e., the funds the investor deposited in the market), is independent of the initial endowment.

3.1.3 Rankings

The market we study provides ranking information of investors based on their PnL and RoI. The investor with the highest PnL (or RoI) ranks first, and other investors are sorted in descending order. Crucially, this ranking includes inactive investors who registered on the market but do not have any positions. These inactive investors have, by definition, a PnL and a RoI of zero, which is higher than that of investors who have incurred losses. As a result, the rank of an investor with a slightly negative PnL/RoI is orders of magnitude larger than that of an investor with a slightly positive PnL/RoI.

3.1.4 Cryptocurrency prices

The exchange also provides real-time prices of major cryptocurrencies via its public API. We collect these prices every minute throughout our measurement period. All collected prices are denominated in Tether (USDT).

3.2 Data collected

The cryptocurrency derivatives exchange we study started to publish ranking information on a leaderboard in mid-2020. While the leaderboard web front-end only shows the top investors, the public API initially provided information on every investor on the platform. Ranking data was updated hourly until May 9, 2021. Updates then shifted to a daily basis, until July 26, 2021. At that point, the exchange stopped providing ranking data for all investors; instead, the API now merely matches what the web front-end shows. As a result, we use data collected between August 20, 2020 and July 20, 2021.

4 Estimating the number of investors

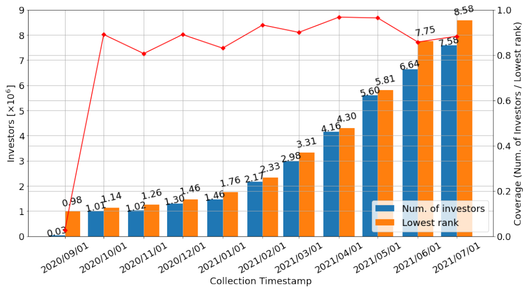
4.1 Number of investors

As discussed above, the exchange API provides performance indices and ranking data about all investors. Unfortunately, to query data about a specific investor, we need their ID, and we cannot directly obtain the number of investors on the platform. Instead, we use ranking data as a proxy to estimate it.

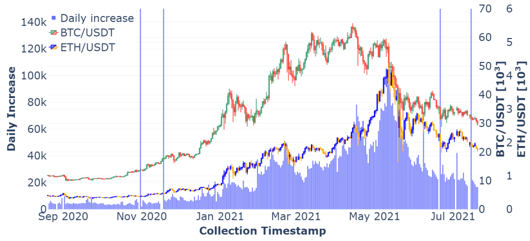
Figure 1 shows the number of investors in our dataset, the maximum PnL rank among the investors, and their ratio at the beginning of each month in our observation period.

The figure shows that we collected data on more than one million investors and this ratio stays above 0.80 after October 2020—the first month is an anomaly due to our data covering only a week or so. The large sample size ensures the lowest rank among collected investors is statistically very close to the number of investors in the market.[5] With this in

[5] As a rough estimation, the probability that the relative error between the lowest rank and the (actual) number of investors in the market is equal to or less than 0.001% throughout our observation period (293 days) with one million samples ($\sim$ the number of investors at the beginning of October 2020) is: $Pr(\text{Relative Error} < 0.001\%) = \left(1 - (1 - 0.00001)^{1{,}000{,}000}\right)^{293} \simeq 0.987$. Given the increasing sample size, the actual probability is better than the approximation.



■ **Figure 1** The number of investors in our dataset and the maximum (lowest) rank among the investors



■ **Figure 2** The daily increase in the number of investors in the market and the prices of Bitcoin (BTC) and Ether (ETH).

mind, Figure 1 shows 7.5 million new investors joined the market increased in the ten months between September 1, 2020 to July 1, 2021.

Using maximum PnL rank as a proxy, we can estimate the number of investors in the market on a daily basis, even with an imperfect coverage of investors. Figure 2 shows both the daily increase in users on the platform, and the Bitcoin (BTC) and Ethereum (ETH) spot prices. Graphically, there seems to be a strong correlation between the number of new users joining the market, and the price of these currencies. The outliers (abnormally large increases) in November 2020 and July 2021 come from data collection errors due to changes in the exchange API implementation and collector breakdown.

In Section 6, we refine this intuition with a complete regression analysis.

4.2 Leaderboard data idiosyncracies

We have to account for certain idiosyncracies in our data. We infer registration numbers from the leaderboard data, which we itself get from a public API. However, there may be some lag times between what the API returns (leaderboard data may not be faithfully updated in real-time), and actual numbers; this can have an impact on our regression analysis.

Figure 3 shows when new user registrations appear in our data, on a hourly basis. Each point corresponds to the relative increase in number of registered users compared to the previous hour, using the maximum leaderboard rank among observations in the hour as a proxy, as discussed earlier. We plot this data over our complete measurement interval (so, roughly 7,000 points corresponding to the number of hourly samples in our 10-month data). We observe that the reported number of users jumps during 0:00-3:00AM UTC on most days and usually does not change much thereafter. From this behavior, we hypothesize that the exchange updates the set of investors in the performance rankings once a day at midnight, integrating most, if not all, of those who registered in the previous day at that time.

Therefore, we define the number of investors in the market in a day d as $I_d \equiv \max_{\tau \in d+1} I_\tau$, where I_τ is the largest observed leaderboard rank in a time slice τ. We also define the daily increase in a day d (N_d) as $N_d \equiv I_d - I_{d-1}$.

5 Regression analysis

We start by discussing the regression variables, before exploring how to construct our regression, considering the properties of the data we have at our disposal.

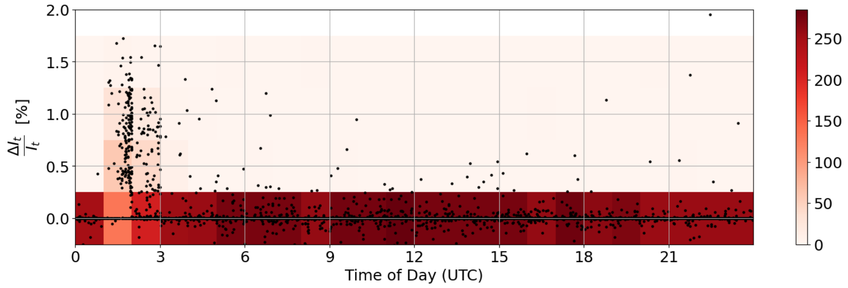


Figure 3 Hourly relative increase in the number of investors. The black dots show the exact time the largest rank in an hour was observed and the relative increase from the previous hour's largest rank. The background color shows the number of observations for a block of an hour and a 0.5% relative increase.

5.1 Variables

5.1.1 Daily user increase

Our first month of data has problematically sparse samples (3.0×10^4) and low coverage (2.67%). Hence, we discard it, and limit our analysis to October 1, 2020–July 20, 2021. We fix the handful of discontinuities observed in Figure 2 – due to data collection errors – by removing the outliers and replacing them with linear interpolations.

As Figure 2 shows, the daily increase N_d does not converge or revert to a mean value. In fact, as we will see in Section 6, N_d is a non-stationary variable. Fortunately, the Box-Cox transformation [9, 52] allows us to include such variables in an autoregressive model like the one we consider, by instead using a transformed variable that satisfies certain properties.[6] In our case, the logarithm of the daily increase, $\log N_d$, satisfies these requirements.

5.1.2 Prices

As noted above, we gather per-minute cryptocurrency prices. For currency X, at day d, we thus collect a vector of prices $\mathbf{P_{X,d}} = \{P_{X,1}, \ldots, P_{X,1440}\}$ corresponding to the 1 440 minutes in a day. The realized daily volatility $\sigma_{X,d}$ is:

$$\sigma_{X,d} = \sqrt{\frac{1440}{|\mathbf{P_{X,d}}|} \sum_{\tau \in d, \tau > 1} \left(\log P_{X,\tau} - \log P_{X,\tau-1}\right)^2} \ ,$$

where $P_{X,\tau}$ is the price of cryptocurrency X measured at time τ in day d.

Here too we use a Box-Cox transformation, and consider the logarithm of the daily average prices, $\log \bar{P}_{X,d}$, as an explanatory variable. Its first difference $\Delta \log \bar{P}_{X,d} \equiv \log \bar{P}_{X,d} - \log \bar{P}_{X,d-1}$ is the logarithmic return of the price, showing the approximate percentage change in the daily price. We will also use the realized volatility $\sigma_{X,d}$ as an additional explanatory variable. To calculate daily average prices $\bar{P}_{X,d}$ in a manner robust to short-lived volatile price movements, we will follow Biais et al. [7], by calculating the average of median values over short time intervals (5 minutes).

We select four cryptocurrencies for their importance and/or unique characteristics.

Bitcoin (BTC): Bitcoin has the largest market cap among cryptocurrencies, and is frequently touted as the "reserve currency" of the cryptocurrency ecosystem. BTC-USDT is the most

[6] Namely, that the mean and variance of its first difference are stationary.

popular futures contract in the exchange we consider, and Bitcoin presents the largest open interest, that is, the total amount (in USDT) of futures contracts held by market participants.

Ethereum (ETH): Ethereum has the second largest market cap among cryptocurrencies, and features the second largest open interest in the exchange. ETH is the utility token in the Ethereum blockchain, which supports many smart contracts, including the majority of decentralized finance (DeFi) contracts and protocols. ETH thus gives us some insights into potential investor interests (and beliefs) in more elaborate blockchain proposals.

Ripple (XRP): XRP is another major cryptocurrency with a decentralized consensus mechanism [10]. Ripple Labs, Inc., the company behind XRP, was sued by the U.S. Securities and Exchange Commission (SEC) in December 2020.[7] At the time of writing, the suit has not been resolved. Among all cryptocurrency legal wranglings, this case is interesting to understand the potential influence of regulatory measures on user interest in a pretty popular coin, specifically, the third largest coin by market capitalization at the time.[8] Hence, XRP could give us insight into investor reactions to regulatory issues.

Dogecoin (DOGE): Originally a "meme" cryptocurrency primarily designed with humorous goals in mind, DOGE received increased attention due to numerous social media campaigns by influencers touting its potential (notably for tips and micropayments). As a result of the attention, DOGE soared in value from 0.005 USDT in January 2021 to 0.5 USDT in May 2021, before hitting an all-time high of 0.75 USDT on May 7, 2021. Social media attention faded away shortly thereafter, and the currency lost significant value. DOGE is thus an interesting currency to include, as a loose proxy for social media activity.

Table 1 summarizes statistics for the logarithms and realized volatilities for the four cryptocurrencies above. Reflecting the price hike in DOGE in early 2021, the standard deviations for DOGE are higher than other variables. We will later use the mean values and standard deviations of level variables for the Principal Component Analysis (PCA). Appendix A shows the plot of daily average prices and realized volatilities of four selected cryptocurrencies.

Table 1 Descriptive statistics for the daily increase in the number of investors and the logarithm of daily average price and realized volatility of BTC, ETH, XRP, and DOGE

	$\log N$	$\log \bar{P}_{BTC}$	$\log \bar{P}_{ETH}$	$\log \bar{P}_{XRP}$	$\log \bar{P}_{DOGE}$	σ_{BTC}	σ_{ETH}	σ_{XRP}	σ_{DOGE}
• **Level variable**									
Mean	9.862	10.371	7.145	-0.636	-3.396	0.047	0.059	0.087	0.100
Median	10.077	10.469	7.434	-0.610	-2.917	0.042	0.051	0.071	0.066
Std. Dev.	0.899	0.518	0.714	0.583	1.923	0.025	0.038	0.060	0.099
Max.	11.465	11.059	8.346	0.589	-0.370	0.233	0.475	0.417	0.900
Min.	7.654	9.261	5.827	-1.556	-5.990	0.010	0.019	0.018	0.015
• **First difference**									
Mean	0.006	0.004	0.006	0.003	0.014				
Median	-0.012	0.006	0.006	0.002	0.001				
Std. Dev.	0.212	0.037	0.046	0.077	0.127				
Max.	1.174	0.126	0.163	0.291	1.238				
Min.	-1.256	-0.145	-0.196	-0.321	-0.470				

[7] See `https://www.sec.gov/news/press-release/2020-338`.
[8] See `https://coinmarketcap.com/historical/20201220/`

5.2 Method

All of our variables are time-dependent and potentially highly correlated. An unbiased regression analysis generally requires time-dependent variables to be at least (weak-)stationary [4, 19, 20, 30] and to present low correlation [39]. Stationarity means the mean values should be finite, time-invariant, and auto-covariances should only depend on the time interval over which they are calculated. By successively differencing a non-stationary variable, we might eventually end up with a stationary variable (e.g., a random walk variable y_t following $y_t = y_{t-1} + \epsilon_t$ with white noise ϵ_t is not stationary, but its first difference, $\Delta y_t = \epsilon_t$, is). We denote by $I(d)$ the number of successive differencing operations required to make the tested variable stationary. $I(d)$, also called the *order of integration*, will be key in determining which regression model to use. Also, keeping the correlation between explanatory variables low is an essential part of pre-processing to hold a regression analysis informative.

5.2.1 Unit root test

To check stationarity, we rely on the *unit root test* technique. One of the best known such tests is the Augmented Dickey-Fuller (ADF) test [14]. ADF tests the null hypothesis that the variable tested is a unit root (i.e., $I(1)$). If it rejects the null hypothesis with a small enough p-value, the process is deemed stationary ($I(0)$). The Phillips-Perron (PP) test [38] is also widely used to test stationarity. PP assumes the same null hypothesis as ADF, but allows heteroskedasticity and autocorrelation in the error term. We will use both PP and ADF in our analysis.

5.2.2 Principal Component Analysis

We employ *Principal Component Analysis* (PCA, [18]) to solve the problem of high correlation between explanatory variables. PCA is an orthogonal projection of the original variables (X) onto a lower-dimensional set of variables (S_L), preserving as much information as possible: $S_L = \widehat{X} W_L$, where L is the dimension of PCA-vector space ($L \leq \dim(X)$). W_L is the coefficient matrix for constructing principal components from normalized price-related variables $\widehat{X_i}$, which is composed of the variables normalized with its mean value ($\overline{X_i}$) and standard deviation ($\sqrt{Var(X_i)}$): $\widehat{X_i} \equiv \frac{X_i - \overline{X_i}}{\sqrt{Var(X_i)}}$. Because PCA components are orthogonal, PCA prevents the regression analysis from being contaminated by highly correlated components. We can then calculate the original variables' coefficients from those for PCA components by simple linear algebraic manipulations.

5.2.3 Autoregressive distributed lag model

We will build our regression using an autoregressive distributed lag (ARDL) model, which, contrary to most regression models, can accommodate a mixture of $I(0)$ variables and $I(1)$ variables [36]. This makes it particularly suited to our problem, given the apparent non-stationarity of at least some of our variables.

We will use the following unrestricted error correction model (UECM) representation of

ARDL in our analysis:

$$\begin{aligned}\Delta\widehat{\log N_d} \quad = \quad & c_0 + \textstyle\sum_S \gamma_S I_{S,d} \\ & +\pi_0 \widehat{\log N_{d-1}} + \textstyle\sum_i \pi_i v_{i,d-1} + \sum_i \pi_i' w_{i,d-1} \\ & + \textstyle\sum_{i=1}^{p-1} \alpha_i \Delta \widehat{\log N_{d-i}} \\ & + \textstyle\sum_i \sum_{j=0}^{q_i-1} \beta_{i,j} \Delta v_{i,d-j} + \sum_i \sum_{j=0}^{q_i'-1} \beta_{i,j}' \Delta w_{i,d-j} \\ & +\epsilon_d \ ,\end{aligned} \tag{1}$$

where p, $q_i(q_i')$, $I_{S,d}$, and ϵ_d represent the lag order of the normalized daily increase $\left(\widehat{\log N_d} \equiv \frac{\log N_d - \overline{\log N_d}}{\sqrt{Var(\log N_d)}}\right)$, those for principal components for daily average prices (v_i) and realized volatilities (w_i) in d-th day, indicator variables of interest (labeled by S), and the error term, respectively. α, β, β', γ, π_0, π, and π' are regression coefficients.

Pesaran et al. [36] propose a bounds test in an ARDL model (*PSS-bounds test*), to determine the existence of a long-run equilibrium relationship (i.e., cointegration) between variables. The test compares the test statistic with two critical boundaries. If the tested statistic is larger than the upper boundary (called $I(1)$-boundary), the test confirms the existence of a long-run relationship; On the other hand, if the tested statistic is lower than the lower boundary ($I(0)$-boundary), the test rejects the existence of a long-run relationship. If the tested statistic falls between the $I(0)$ and $I(1)$ boundary, no conclusion about the existence, or lack thereof, of a long-run relationship can be derived. PSS-bounds test has five cases (Case I-V) for the specification of deterministic terms. We consider Case I (no constant term in the ARDL model), Case II (a constant term in the ARDL model and cointegration), and III (a constant term in the ARDL model, but no constant term in cointegration). In the UECM representation, the cointegrations are mainly given by the second line in Eqn. (1):

$$\begin{aligned} \widehat{\log N_d} + \tfrac{1}{\pi_0}\left(\textstyle\sum_i \pi_i v_{i,d} + \sum_j \pi_j' w_{i,d}\right) \quad &= \quad 0 \quad \text{(Case I)} \ , \\ \widehat{\log N_d} + \tfrac{1}{\pi_0}\left(\mu + \textstyle\sum_i \pi_i v_{i,d} + \sum_j \pi_j' w_{i,d}\right) \quad &= \quad 0 \quad \text{(Case II)} \ , \\ \widehat{\log N_d} + \tfrac{1}{\pi_0}\left(\textstyle\sum_i \pi_i v_{i,d} + \sum_j \pi_j' w_{i,d}\right) \quad &= \quad 0 \quad \text{(Case III)} \ , \end{aligned} \tag{2}$$

where μ is the deterministic term(s) for cointegration.

Intuitively, Eqn. (1) says that the change in $\widehat{\log N_d}$ is explained by (1) the short-run change in itself and explanatory variables and (2) the deviation from cointegration (i.e., long-run equilibrium status) if it exists.

We can consider the marginal effect of explanatory variables ($\frac{\partial \widehat{\log N_{d+k}}}{\partial V_{X,d}}$) in an arbitrary temporal duration k (≥ 0) when they converge to zero over time.

Short-run multipliers ($\frac{\partial \widehat{\log N_d}}{\partial V_{X,d}}$) represents the immediate impact of an explanatory variable $V_{X,d}$. In Eqn. (1), short-run multipliers are given by $\beta_{i,0}$ and $\beta_{i,0}'$. The cumulative marginal effect up to k-th day ($\sum_{l=0}^{k} \frac{\partial \widehat{\log N_{d+k}}}{\partial V_{X,d+l}}$) shows the accumulated impact of change in explanatory variables lasting for k days, and converges to a finite value as k increase when the marginal effect converges to zero. Since $\sum_{l=0}^{k} \frac{\partial \widehat{\log N_{d+k}}}{\partial V_{X,d+l}} = \sum_{l=0}^{k} \frac{\partial \widehat{\log N_{d+l}}}{\partial V_{X,d}}$, we can also interpret this quantity as the cumulative effect that today's change in an explanatory variable will cause for k days in the future.

Long-run multipliers ($\lim_{k\to\infty} \sum_{l=0}^{k} \frac{\widehat{\log N_{d+k}}}{\partial V_{X,d+l}}$) denote the cumulative marginal effect on $\widehat{\log N_{d+k}}$ coming from a persistent change in an explanatory variable. From the discussion above, this quantity represents the cumulative effect today's change in an explanatory variable causes in the long future. Going back to Eqn. (1), long-run multipliers are given by $\frac{-\pi}{\pi_0}$ and $\frac{-\pi'}{\pi_0}$ for the principal components of daily average prices and realized volatilities, respectively.

Our analysis considers two major regulatory measures that affected cryptocurrency prices in our observation period, using two indicator variables (I_S): (1) the influence of the SEC litigation against Ripple Labs, Inc. and (2) the Chinese government's statement that it planned to tighten cryptocurrency regulation. The big swings in XRP price after the announcement of the lawsuit may affect newly-participating investor behavior. To capture the potential effects, we introduce the indicator variable:

$$I_{SEC,d} = \begin{cases} 1 & d \text{ is before Dec. 22, 2020} , \\ 0 & \text{otherwise} . \end{cases} \tag{3}$$

Dec. 22, 2020 is the day the SEC publicly announced the lawsuit. In the definition of Eqn. (3), the sum of the constant term and I_{SEC} ($c_0 + I_{SEC}$) reresents the constant percentage change in user registrations before the lawsuit was announced; this becomes a constant term (c_0) after that announcement. We employ this definition of I_{SEC} to avoid shifting the critical values of the PSS-bounds test [36].[9]

We use another indicator variable to capture the effect of the Chinese government's statement. It was published on May 21, 2021 [5]. This statement is considered to have had a significant impact on wide range of cryptocurrencies adversely.

$$I_{CHN,d} = \begin{cases} 1 & d \geq \text{May 21, 2021} , \\ 0 & \text{otherwise} . \end{cases} \tag{4}$$

A statistically significant coefficient for I_{SEC} and I_{CHN} would indicate a spill-over effect that is not absorbed in cryptocurrency prices. Geofencing has been an issue for major crypto-exchanges as evidenced by multiple legal proceedings [46–48, 51], with investors allegedly residing in countries that restrict participation (specifically, the US and China). We have no reason to believe the market we study is immune to geofencing issues. Hence, we expect the announcement of these regulatory actions to impact potential investor behavior. Moreover, these measures were announced within our observation period, making it possible to precisely gauge their impact. We also considered the UK ban on retail crypto-derivatives trading that became effective on Jan. 6, 2021[10] as a potentially relevant case, but did not observe any significant impact. We cannot distinguish whether this is because the announcement was made before our observation period started (June 10, 2020), and investors had already factored it into account, or because UK regulations have less of an overall impact.

Our analysis uses *urca* package for R [37] for unit-root tests and *statsmodels* package for Python [42] for the remaining analyses. We employ heteroskedasticity autocorrelation (HAC) robust variance estimation throughout our analyses to compensate for the potential impact of determinants other than our selected terms and autocorrelation.

6 Results

We start with unit root tests to ensure all variables are $I(0)$ or $I(1)$ so that we can use ARDL. Then, we consider the correlation between explanatory variables and finally perform a complete analysis of our ARDL model to tease out the factors behind user registrations.

[9] PSS-bounds test's critical values must be modified if the regression formula includes indicator variable(s) that do not disappear as the observation period increases. We defined I_{SEC} in Eqn. (3) to mitigate the potential contamination from long-lasting non-zero indicator variables.

[10] https://www.fca.org.uk/news/press-releases/fca-bans-sale-crypto-derivatives-retail-consumers

Table 2 Unit root test results.

Variable	Level variable: Intercept	Level variable: Intercept and Trend term	First difference: Intercept	First difference: Intercept and Trend term	Order of integration
ADF test					
$\log N$	-1.91	-0.49	-10.07***	-10.31***	$I(1)$
$\log \bar{P}_{BTC}$	-2.38	-0.45	-6.53***	-7.23***	$I(1)$
$\log \bar{P}_{ETH}$	-1.84	-0.60	−10.52***	−10.70***	$I(1)$
$\log \bar{P}_{XRP}$	-1.71	-1.55	−11.12***	−11.16***	$I(1)$
$\log \bar{P}_{DOGE}$	-1.10	-1.39	−7.85***	−7.87***	$I(1)$
σ_{BTC}	-4.04***	-4.02***	-9.40***	-9.41***	$I(0)$
σ_{ETH}	−4.11***	−4.17***	−9.78***	−9.78***	$I(0)$
σ_{XRP}	−5.31***	−5.29***	−7.70***	−7.71***	$I(0)$
σ_{DOGE}	-6.44***	−6.53***	−9.67***	−9.66***	$I(0)$
PP test					
$\log N$	-1.82	-1.98	-26.53***	−26.93***	$I(1)$
$\log \bar{P}_{BTC}$	-2.39	-0.27	−14.25***	−14.65***	$I(1)$
$\log \bar{P}_{ETH}$	-1.85	-0.44	−13.03***	−13.15***	$I(1)$
$\log \bar{P}_{XRP}$	-1.72	-1.57	−13.03***	−13.04***	$I(1)$
$\log \bar{P}_{DOGE}$	-1.06	-1.29	−14.60***	−14.60***	$I(1)$
σ_{BTC}	−8.18***	−8.40***	−31.31***	−31.30***	$I(0)$
σ_{ETH}	−9.47***	−9.77***	−33.92***	−33.88***	$I(0)$
σ_{XRP}	−7.69***	−7.70***	−26.83***	−26.80***	$I(0)$
σ_{DOGE}	−6.80***	−6.91***	−21.84***	−21.80***	$I(0)$

*, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

6.1 Unit root test

Table 2 summarizes the unit root test results for level variables and their first difference, where we determine the lag order in ADF to minimize Akaike Information Criterion (AIC) [1]. Both ADF and PP provide consistent results about variable stationarity. The analysis shows that (taking their logarithms), the daily user registration increases and the daily average prices are unit root $I(1)$, but volatilities are stationary, i.e., $I(0)$. Hence, we can use ARDL in our analysis for daily registrations and PCA components constructed from price-related variables: the PCA components, which are composed of the linear combination of $\log \bar{P}_X$ and σ_X, are at most $I(1)$.

6.2 Principal Component Analysis

Table 3 shows that the Pearson correlation coefficients between log daily average prices and realized volatilities are so high that regression analysis with these variables will suffer from a multi-collinearity problem [18, 39].

Therefore, we consider the Principal Component Analysis (PCA) of daily average prices and realized volatilities. Table 4 summarizes the construction of PCA components from normalized log daily average prices ($\widehat{\log \bar{P}_X}$) and realized volatilities ($\widehat{\sigma_X}$). The table shows that the first component (PC1) for both log daily average prices (v_1) and realized volatilities (w_1) are composed of the almost equally weighted sum of four coins, which basically denotes the *average* trend of cryptocurrency prices and volatilities. The second component in daily average prices (PC2) has large BTC and XRP coefficients with opposite signs, capturing how XRP price trends deviate (or get "decoupled") from BTC price trends, to which the SEC litigation against Ripple may have contributed. The realized volatilities' second component measures the volatility difference between major coins (BTC and ETH), on the one hand,

Table 3 Pearson correlation coefficients for daily average prices and realized volatilities.

	Daily average price				Realized volatility			
	$\log \bar{P}_{BTC}$	$\log \bar{P}_{ETH}$	$\log \bar{P}_{XRP}$	$\log \bar{P}_{DOGE}$	σ_{BTC}	σ_{ETH}	σ_{XRP}	σ_{DOGE}
$\log \bar{P}_{BTC}$	1.00	0.90	0.64	0.78	0.31	0.25	0.27	0.33
$\log \bar{P}_{ETH}$		1.00	0.78	0.95	0.31	0.30	0.20	0.31
$\log \bar{P}_{XRP}$			1.00	0.81	0.12	0.18	0.25	0.24
$\log \bar{P}_{DOGE}$				1.00	0.25	0.25	0.13	0.29
σ_{BTC}					1.00	0.92	0.56	0.52
σ_{ETH}						1.00	0.57	0.49
σ_{XRP}							1.00	0.54
σ_{DOGE}								1.00

Table 4 Principal component coefficients and percentage of variance explained by each principal component.

	Daily average price					**Realized volatility**				
	$\widehat{\log \bar{P}_{BTC}}$	$\widehat{\log \bar{P}_{ETH}}$	$\widehat{\log \bar{P}_{XRP}}$	$\widehat{\log \bar{P}_{DOGE}}$	% of variance	$\widehat{\sigma_{BTC}}$	$\widehat{\sigma_{ETH}}$	$\widehat{\sigma_{XRP}}$	$\widehat{\sigma_{DOGE}}$	% of variance
PC1	0.015	0.017	0.015	0.016	86.0	0.019	0.019	0.016	0.015	70.4
PC2	-0.060	-0.018	0.071	0.011	9.4	-0.031	-0.033	0.027	0.050	16.3
PC3	-0.074	0.036	-0.069	0.095	4.1	-0.013	-0.007	0.070	-0.050	11.3
PC4	-0.146	0.368	0.008	-0.246	0.4	-0.145	0.144	-0.007	0.009	2.0

and relatively less prominent coins (XRP and DOGE), on the other hand. Figure 4 shows the first and second components for log daily average prices (v_1, v_2) and realized volatilities (w_1, w_2). As we expect from Table 4, the first component for log daily average prices (v_1) represents cryptocurrency price trends: rising until May 2021 and the subsequent downturn. The second component for log daily average (v_2) prices denotes a sudden decrease in the value in late December 2020, when the SEC announced its litigation against Ripple. The increase in early April 2021 might be caused by investors getting more relaxed about the impact of this litigation on XRP [23]. Finally, the second component for realized volatilities (w_2) displays sharp positive spikes in February and April 2021 caused by XRP and DOGE as well as the negative spike in May 2021 due to the high volatility of BTC.

6.3 ARDL model analysis

We next delve into our regression analysis with the ARDL model. We determine the lag order of autoregressive terms ($\Delta \widehat{\log N_d}$) and distributed lag terms ($\Delta v_{i,d}$ and $\Delta w_{i,d}$) to minimize the Bayesian information criterion (BIC) [41]. In determining the lag orders, we limit ourselves to a maximum lag order of ten for both autoregressive terms and distributed lag terms. This means that we consider a lag of up to ten days. Then, we select a model with the smallest BIC from those with lag orders higher than or equal to one for all distributed lag terms, so that we can construct a UECM representation. Fortunately, models with smaller lags yield smaller BIC values than those with higher orders, so our self-imposed limitation for the maximum lag order does not affect our results.

Model Specification: We consider five models, summarized in Table 5, for analyzing the influence of cryptocurrency prices on user registrations to a cryptocurrency derivatives market. Models 1–3 analyze the effect of model complexity. Models 4 and 5 measure the influence of regulatory measures on daily registration by comparing them with Model 3. During model selection, we found that models with different combinations of principal components all reduced to those listed in Table 5. For example, a model selection starting from a model

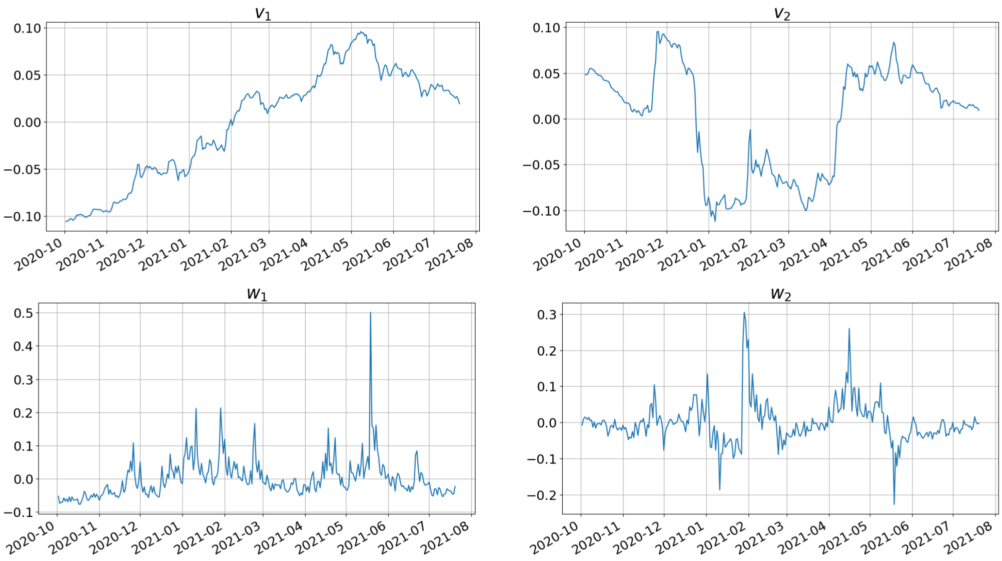


Figure 4 First and second components for log daily average prices (top) and realized volatilities (bottom).

Table 5 Model selection for ARDL analysis

	Num. of PCA components for log daily average prices (Cum. % of variance)	**Num. of PCA components for realized volatilities** (Cum. % of variance)	**Indicator variables**
Model 1	1 (86.0%)	1 (70.3%)	No
Model 2	2 (95.4%)	1 (70.3%)	No
Model 3	2 (95.4%)	2 (86.6%)	No
Model 4	2 (95.4%)	2 (86.6%)	Chinese govt. statement
Model 5	2 (95.4%)	2 (86.6%)	Chinese govt. statement + SEC XRP lawsuit

with the first principal component for log daily prices and the first and second components for realized volatilities reduces to Model 1 in optimization.

6.3.1 Fitting result

Table 6 summarizes the estimation results for all ARDL regressions. First, our full-fledged Model 5 exhibits minimum values for all information criteria. That indicates Model 5 is the best among fitted models. Model 4 presents the second-smallest information criteria values. These results indicate that adding the indicator variables for controlling regulatory measures, as well as the selection of principal components, enhances the explanatory power of our ARDL models.

Second, the first difference of the first principal component (PC1) for the logarithm of the daily average price ($\Delta v_{1,d}$) significantly influences user registrations. Given the standard deviation for the daily registration ($\log N$) and the logarithm of daily average prices ($\log \bar{P}_X$) in Table 1 and the coefficients for PC1 in Table 4, a 1.0% increase in cryptocurrency prices for a given day will roughly drive a 2.0% increase in user registrations in the same day.[11]

[11] Due to normalization while constructing the PCA components, we have to multiply the ratio of standard deviations ($\sqrt{Var(\log N)/Var(\log P_X)}$) and the coefficient for constructing PCA (Table 4) to the

Table 6 ARDL regression results for Models 1–5. The values in parentheses are standard errors.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Const. (c_0)	-0.002	-0.005	-0.004	0.038**	−0.041
	(0.010)	(0.010)	(0.010)	(0.015)	(0.034)
$\log \widehat{N_{d-1}}$	−0.281***	−0.440***	−0.477***	−0.707***	−0.712***
	(0.073)	(0.096)	(0.099)	(0.104)	(0.099)
$\Delta \log \widehat{N_{d-1}}$	−0.234**	−0.152*	−0.145*	–	–
	(0.095)	(0.089)	(0.090)		
Log daily average price					
$v_{1,d-1}$	4.193***	6.749***	7.268***	11.384***	12.730***
	(1.098)	(1.474)	(1.517)	(1.734)	(1.695)
$v_{2,d-1}$	–	−1.050***	−1.211***	−1.401***	−2.371***
		(0.253)	(0.275)	(0.261)	(0.467)
$\Delta v_{1,d}$	22.940***	26.536***	24.339***	23.245***	24.116***
	(2.506)	(2.596)	(3.265)	(2.958)	(2.916)
$\Delta v_{2,d}$	–	−4.764***	−4.783***	−5.091***	−6.388***
		(1.187)	(1.170)	(1.194)	(1.425)
Realized Volatility					
$w_{1,d-1}$	0.960***	1.334***	1.404***	1.990***	2.241***
	(0.344)	(0.336)	(0.360)	(0.308)	(0.341)
$w_{2,d-1}$	–	–	0.588***	0.343**	0.488***
			(0.211)	(0.174)	(0.213)
$\Delta w_{1,d}$	2.017***	2.136***	2.038***	1.900***	2.028***
	(0.450)	(0.390)	(0.328)	(0.349)	(0.347)
$\Delta w_{1,d-1}$	0.793***	0.638***	0.594**	–	–
	(0.220)	(0.216)	(0.235)		
$\Delta w_{2,d}$	–	–	0.138	0.343	0.177
			(0.303)	(0.174)	(0.301)
Indicator variables					
I_{CHN}	–	–	–	−0.201***	−0.145***
				(0.047)	(0.049)
I_{SEC}	–	–	–	–	0.239**
					(0.095)
PSS-bounds test					
Case-I (w/o const.)	5.263***	5.395***	4.729***	7.442***	7.442***
Case-II (w const.)	3.969**	4.411***	4.063**	6.182**	6.182***
Case-III (w/o const.)	5.271**	5.457***	4.779***	7.408***	7.408***
Best-fit model	UECM(2,1,2)	UECM(2,1,1,2)	UECM(2,1,1,2,1)	UECM(1,1,1,1,1)	UECM(1,1,1,1,1)
Num. of observations	292	292	292	292	292
Log-Likelihood	64.647	78.637	82.623	95.690	99.369
AIC	-111.294	-135.274	-139.246	-167.380	-172.737
BIC	-78.265	-94.905	-91.538	-123.300	-124.984
HQIC	-98.061	-119.100	-120.132	-149.721	-153.607
R^2	0.325	0.387	0.404	0.454	0.467

***, **, and * represent significance at the 1%, 5%, and 10% level.

This pattern consistently shows up in all models, which indicates that rising cryptocurrency prices positively correlate with decisions of potential investors to join the market.

Third, the first difference of the first principal component for the realized volatilities ($\Delta w_{1,d}$), i.e., the change in the overall volatility trend, shows a similar influence pattern. The same-day increase in the variable ($\Delta w_{1,d}$) consistently has a significant impact on the daily increase in the number of investors in all models. In the original variables scale, a 0.01 increase in all realized volatilities causes a 3.0% larger user registration on the same day.

These influence patterns of (the logarithm of) daily average prices and realized volatilities are consistent with often heard narratives about motivations for engaging in cryptocurrency investments: cryptocurrency investors are supposedly primarily driven by speculation, so

coefficient for ARDL in Table 6 to get the coefficient in their original scales.

cryptocurrency price rise and high volatilities will drive more user participation.

However, our regression analysis also shows a more complex picture of the factors influencing investor behavior. Model 4, which includes the indicator variable that captures the potential impact of the Chinese government's statement (I_{CHN}), suggests that the constant term (c_0) is positive and significant at the 5% level. This implies that the daily registration increases ($\log N$) by 3.4% every day in the original scale, which is given by multiplying I_{CHN} by the standard deviation of $\log N$ ($= 0.038 \times 0.899$), even if cryptocurrency prices were stable *before* the statement was published. However, our analysis shows that the Chinese government statement poured cold water on investor enthusiasm. Specifically, the influence of I_{CHN} term swallows the constant term, and the sum of these two terms ($c_0 + I_{CHN}$) turns to negative (-0.163), meaning that new registrations will decrease by 14.7% ($= 0.163 \times 0.899$) every day in the original scale if cryptocurrency prices are stable. This result evidences the strong impact of a specific regulatory issue on investor behavior that is not explained by decreasing cryptocurrency prices. Note that the constant term for Model 4 does not have to be zero, although we employ PCA for both the dependent and explanatory variables. This is because the indicator variable (I_{CHN}) is not centered.[12]

Finally, we consider the effect of the SEC litigation against Ripple on user registrations. The constant term (c_0) for Model 5 loses significance at the 5% level, and I_{SEC} holds a large coefficient of 0.239. So, the constant percentage change in user registrations before the lawsuit announcement ($c_0 + I_{SEC}$) is 0.198, suggesting a 17.8% daily increase in user registration in its original scale. However, this increase subsided after the litigation was announced, once again showing that a regulatory issue impacted user behavior.

PSS bounds test result: Next, we consider the long-run effect of prices in detail. Since marginal effects $\left(\frac{\partial \widehat{\log N_{d+k}}}{\partial V_{X,d}}\right)$ for all explanatory variables converge to zero as time goes on (see Appendix B), we can consider a stable long-run equilibrium state.

Since we use normalized variables for regression (see Section 5), the constant terms for Models 1–3 are theoretically zero, consistent with the results in Table 6. Hence, the appropriate bound test case specification for Models 1–3 is Case-I in Eqn. (2). On the other hand, Table 6 shows that the constant term for Model 4 is non-zero at the 1% significance level, indicating that Case-II or Case-III are appropriate. There is no theoretical restriction to determine the appropriate bound test case specification for Model 5, so we consider Case I–III.

Fortunately, all PSS bounds test results in Table 6 reject the null hypothesis that there is no cointegration (i.e., an equilibrium state) between the daily user registration and the price-related variables at the 5% significance level. This result strongly suggests the existence of a long-run equilibrium relationship between the inflow of new investors to the cryptocurrency investment market and cryptocurrency prices.

Figure 5 shows the observed user registration and the estimation from our cointegrations in Models 3–5. It demonstrates that our cointegration replicates the observed data well. This result has crucial implications. Since cryptocurrency derivatives are traded on off-chain exchanges, investor demographics, such as population, are not fully observable. This can cause considerable information asymmetry between market operators and outsiders, such as investors and financial regulators. However, our cointegration may be useful as an easy

[12] A linear regression of a normalized dependent variable with normalized explanatory variables requires that the constant term be zero, as is the case in Models 1–3. However, the sum of the constant term and the average value of the indicator variable(s) has to be zero when the un-centered variable(s) is/are integrated. In Model 4, that sum is $0.038 - 0.201 \times \frac{61}{292} \simeq -4.0 \times 10^{-3}$, which satisfies this condition.

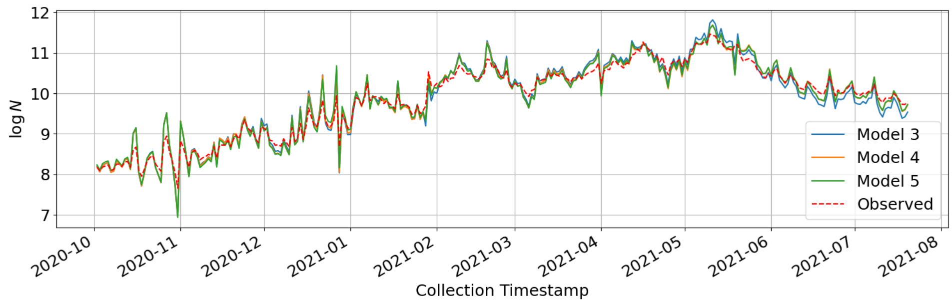


Figure 5 The logarithm of daily user registration increase replicated from our cointegration (solid lines) in Models 3–5 and its observed value (red dashed line)

way to estimate the number of market investors from publicly available price data, thereby reducing this information discrepancy.

6.3.2 Individual cryptocurrency influence

This section considers the influence of each cryptocurrency on daily user registrations. Since a linear algebraic relation connects the original price-related variables and principal components ($S_L = \widehat{X}W_L$), we can derive the coefficients for the daily average prices and realized volatilities in their original scale from those for principal components.

Table 7 summarizes the short-run and long-run multipliers for the daily average prices and realized volatilities in Models 3–5 in their original scales.

Table 7 Long-run and short-run multipliers. The value in the parentheses are standard errors for estimates.

	Model 3		Model 4		Model 5	
Multipliers	Short-run	Long-run	Short-run	Long-run	Short-run	Long-run
$\log \bar{P}_{BTC}$	1.145^{***}	0.668^{***}	1.148^{***}	0.633^{***}	1.307^{***}	0.822^{***}
	(0.181)	(0.035)	(0.176)	(0.028)	(0.194)	(0.082)
$\log \bar{P}_{ETH}$	0.621^{***}	0.378^{***}	0.605^{***}	0.384^{***}	0.653^{***}	0.452^{***}
	(0.085)	(0.011)	(0.079)	(0.008)	(0.080)	(0.028)
$\log \bar{P}_{XRP}$	0.034	0.071^{*}	-0.025	0.151^{***}	-0.146	0.045
	(0.112)	(0.037)	(0.116)	(0.031)	(0.302)	(0.054)
$\log \bar{P}_{DOGE}$	0.161^{**}	0.103^{***}	0.151^{***}	0.112^{***}	0.151^{***}	0.119^{***}
	(0.023)	(0.003)	(0.020)	(0.003)	(0.020)	(0.004)
σ_{BTC}	1.229^{***}	0.638	1.175^{***}	1.373^{***}	1.178^{**}	1.379^{***}
	(0.441)	(0.499)	(0.447)	(0.358)	(0.444)	(0.360)
σ_{ETH}	0.806^{***}	0.352	0.773^{**}	0.881^{***}	0.771^{**}	0.874^{***}
	(0.308)	(0.349)	(0.311)	(0.248)	(0.309)	(0.251)
σ_{XRP}	0.553^{***}	1.211^{***}	0.504^{***}	0.882^{***}	0.566^{***}	1.043^{***}
	(0.129)	(0.199)	(0.116)	(0.113)	(0.132)	(0.155)
σ_{DOGE}	0.346^{**}	0.971^{***}	0.310^{**}	0.613^{***}	0.363^{***}	0.750^{***}
	(0.135)	(0.195)	(0.121)	(0.111)	(0.136)	(0.147)
Const. (μ)	-	0.007	-	-0.054^{***}	-	0.057
		(0.020)		(0.019)		(.050)
C_{ECT} ($-\pi_0$)	-	0.477^{***}	-	0.707^{***}	-	0.712^{***}
		(0.099)		(0.104)		(0.099)

***, **, and * represent significance at the 1%, 5%, and 10% level, respectively.

6.3.2.1 Short-run multipliers

First, we consider the short-run multipliers ($\frac{\partial \log N_d}{\partial V_{X,d}}$), the immediate response of daily registration ($\log N_d$) to the change in an explanatory variable ($V_{X,d}$). Table 7 clearly shows that Bitcoin's average daily price increase and realized volatility have the largest immediate impact on daily user registrations. This result is consistent with the prevailing belief that Bitcoin drives cryptocurrency investments. In fact, a 1.0% increase in the daily average BTC price will cause a 1.1–1.3% increase in user registrations on the same day, and higher volatility can drive registrations further up.

On the other hand, Ripple (XRP) and Dogecoin (DOGE) show smaller immediate impacts from daily average prices and realized volatilities. DOGE appreciation shows positive correlations with user registrations in Models 3–5, but the effect magnitude is roughly one-tenth that of the BTC price. XRP's price changes do not seem to have a significant effect on user registrations.

A potential explanation for these sharp differences across cryptocurrencies lies in their respective popularity. Price swings in Bitcoin and Ethereum gain a lot more media exposure than other cryptocurrencies, which explains the much stronger correlation between the price of these currencies, and the changes in user registrations. On the other hand, although Dogecoin's social media popularity skyrocketed in early 2021, we do not observe a strong direct immediate impact on user registrations; presumably, because this popularity did not immediately percolate to more mainstream media.

6.3.2.2 Long-run multipliers

Next, we consider the long-run multipliers for each cryptocurrency ($\lim_{k\to\infty}\sum_{l=0}^{k}\frac{\partial \log N_{d+k}}{\partial V_{X,d+l}}$), the cumulative influence of the persistent change in an explanatory variable (V_X) on the daily registration ($\log N$). They show an interesting contrast to short-run multipliers.

First, we can observe, in Model 5, a reduction in the long-run multiplier for XRP's daily average price when controlling for the SEC Ripple litigation. In Models 3 and 4, where the indicator variable I_{SEC} is absent, the long-run multiplier is 0.071 (p-value $= 0.053$) and 0.151 (p-value $\simeq 0.000$), indicating the influence is either significant (Model 4), or very close to being significant at the 5% level (Model 3). However, the long-run multiplier for XRP is insignificant even at the 10% level in Model 5. This result, combined with insignificant short-run multipliers, indicates that the XRP price trends lost any importance as a potential investor decision criterion, after the SEC litigation was publicly announced. That is, potential investors basically stopped considering XRP prices when thinking about whether they should join in the derivatives market. Incidentally, this litigation is still proceeding at the time of writing, and is not expected to be resolved between Q3 2023 at the earliest; whether new investors are still ignoring XRP prices in their decision-making, or whether the situation has reverted to what it was before the public announcement of the suit is an interesting open question. (v_2 in Figure 4 hints at a possible return to a state of affairs similar to that before the SEC litigation.)

Regarding realized volatilities, the long-run multipliers show that XRP and DOGE have larger values than BTC and ETH in Model 3. However, in Models 4 and 5, BTC shows the largest impact in both daily average price and realized volatilities, indicating the importance of BTC price also with respect to long-term effects. This implies that not explicitly including the effects of regulatory measures (especially the one in May) would result in a large estimate of the impact of less prominent coins.

6.3.2.3 Cumulative marginal effect

Figure 6 shows the cumulative marginal effect ($\sum_{l=0}^{k}\frac{\partial \log N_{d+k}}{\partial V_{X,d+l}}$) of the daily average prices and realized volatilities in Models 4 and 5. As we discussed in Section 5, the cumulative marginal effect can be interpreted in two ways. First, it denotes the cumulative marginal effect of the change in an explanatory variable (V_X) lasting k days. It also shows the cumulative effect of the change in an explanatory variable that happens today over the future k days (because $\sum_{l=0}^{k} \frac{\partial \log N_{d+k}}{\partial V_{d+l}} = \sum_{l=0}^{k} \frac{\partial \log N_{d+l}}{\partial V_d}$). The result for daily average prices

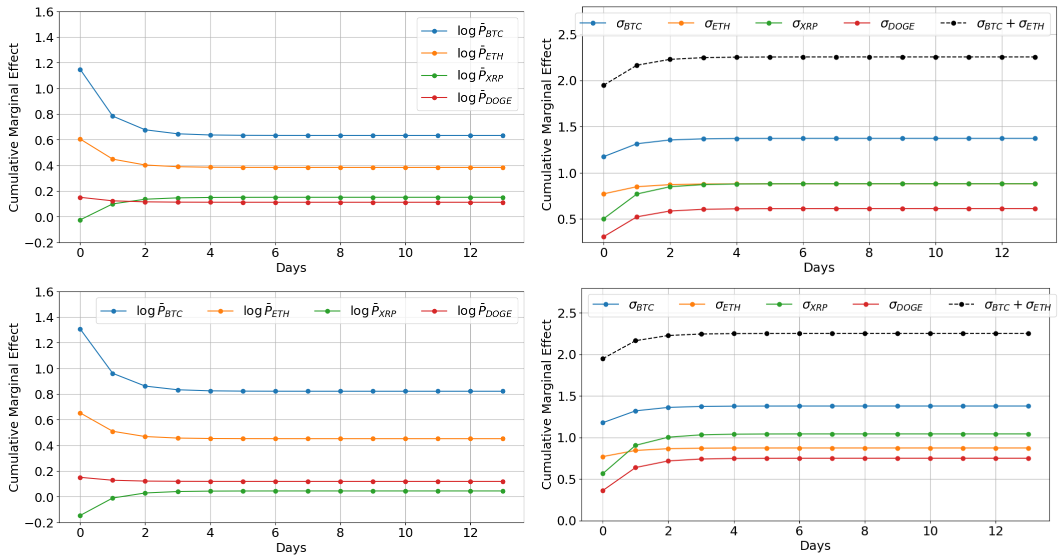


Figure 6 The cumulative marginal effect of the daily average prices and realized volatilities up to two weeks in Models 4 (top panels) and 5 (bottom panels), The black dashed line shows the sum of cumulative marginal effects for BTC and ETH.

shows that the effect of price change peaks immediately; the maximum influence comes on the day the price rises except for XRP (whose prices, as discussed above, do not have a significant short-term impact), and the cumulative effects plateau soon thereafter. In short, user registration increases by a lot immediately, and, then, the positive influence gradually decreases.

The effects of the realized volatility also peak within a few days. However, contrary to the decreasing trend in daily average prices, the cumulative effects pile up as time goes on. The cumulative effects in major coins, BTC and ETH, have a relatively slight gradient since the largest impacts manifest themselves on the same day. This means potential investors immediately react to a volatile situation. Given the chained volatility increase (*volatility clustering*) between BTC and ETH (and others) documented in several pieces of literature [24, 25] – in short, volatility of major coins foster volatile conditions for less prominent currencies as well – the sum of the influences of these coins (black dashed line in Figure 6) seemingly has a measurable market impact. In contrast, the cumulative effects of XRP and DOGE's realized volatilities accumulate by a large number on the next day and the day after that. In short, it takes a longer time for novice crypto investors to digest a volatile situation for relatively minor coins. This is an unsurprising result: contrary to high volatility in BTC and ETH prices, which can attract high publicity in both traditional media and social media, high volatility in less prominent coins, such as XRP and DOGE, will attract the attention of fewer people, which in turn will make its immediate effect more muted. For instance, as noted above, Dogecoin became a social media darling in early 2021, but it took a while for this excitement to propagate to mainstream media, and drive outside investors into cryptocurrency trading.

7 Conclusion

From ranking data on the performance of more than eight million investors in a major cryptocurrency derivatives exchange, we estimated the evolution of the number of market participants from October 1, 2020 to July 20, 2021.

We graphically observed that the daily increase in the number of users seemed to exhibit a strong correlation with major cryptocurrency prices. We formalized this result using the high descriptive capabilities of the autoregressive distributed lag (ARDL) model with Principal Component Analysis (PCA), which accounts for the idiosyncrasies of our data—numerous explanatory variables are not stationary, and are highly correlated.

We empirically analyzed the relationship between the daily user registrations and metrics related to four major cryptocurrencies, Bitcoin, Ethereum, Ripple, and Dogecoin. First, we showed evidence of a long-run equilibrium relationship between the daily registration increase and the prices of the selected cryptocurrencies. The relation is useful for estimating the number of cryptocurrency investors from publicly available price data.

Second, our analysis shows the significant influence of cryptocurrency prices on investor behavior. High price increases and volatility, in general, have the largest impact on user registration on the same day. Among the selected cryptocurrencies, the daily average price of Bitcoin is the largest contributor; this is unsurprising given Bitcoin's leading status among cryptocurrencies. Ethereum prices also significantly impact the daily user registration. In contrast, our analysis shows that Dogecoin prices have a significant but relatively small influence on user registration. A striking result of our analysis is that the impact of Ripple price fluctuations disappears when we control for the SEC litigation against Ripple Labs, Inc. Also, our regression suggests that this lawsuit, and the Chinese government's statements on tightening regulation on cryptocurrency mining and trading have a significant negative impact on user registration. These results indicate the powerful influence of regulatory measures on investor behavior.

Our regression analysis also evidences the impact of price volatility. All coins we selected show significant short-run and long-run effects of volatility on user registrations. This result is consistent with a common narrative that speculation is the primary reason for investors to start investing, so high volatility will attract more people to cryptocurrency exchanges.

However, our analysis also paints a more nuanced picture of the impact of volatility. Volatility effects considerably accumulate over time for relatively minor coins, while they are much more immediate for major cryptocurrencies. This hints at differences in information propagation speed: prominent coins are constantly scrutinized and trends are publicized in real-time, while news updates about less prominent coins initially only reach smaller circles of enthusiasts, mostly on social media, before eventually percolating to the mainstream.

As a limitation, we did not comprehensively assess regulatory measures taken in jurisdictions besides the USA and China. Investor reactions may differ depending on coin specifics, regulation relevance, and jurisdictional importance to exchanges and derivatives trading. However, while limited, our analysis clearly documents examples of the critical influence regulators can have on investor behavior.

Overall, our analysis paints a far more nuanced picture than the simplistic narrative that cryptocurrency derivatives are purely fueled by short-term speculation. Our empirical analysis instead shows potentially complex relationships between prices, volatility, and other factors such as regulatory issues. We hope this could be a starting point to help better understand investors (especially individuals) decisions to participate in cryptocurrency derivative markets, despite the odds being frequently stacked against smaller participants [44].

References

1 Hirotogu Akaike. Information Theory and an Extension of the Maximum Likelihood Principle, page 199–213. Springer Series in Statistics. Springer, 1998.

2 Carol Alexander, Jaehyuk Choi, Hamish R. A. Massie, and Sungbin Sohn. Price discovery and microstructure in ether spot and derivative markets. International Review of Financial Analysis, 71:101506, 2020. URL: https://www.sciencedirect.com/science/article/pii/S1057521920301502, doi:10.1016/j.irfa.2020.101506.

3 Carol Alexander, Jaehyuk Choi, Heungju Park, and Sungbin Sohn. Bitmex bitcoin derivatives: Price discovery, informational efficiency, and hedging effectiveness. Journal of Futures Markets, 40(1):23–43, 2020. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/fut.22050, doi:10.1002/fut.22050.

4 Donald W. K. Andrews. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory, 4(3):458–467, 1988. doi:10.1017/S0266466600013396.

5 Caitlin Ostroff Areddy and James T. Bitcoin, ether prices continue falling after china spurs regulatory fears. Wall Street Journal, May 2021. URL: https://www.wsj.com/articles/bitcoin-ether-prices-continue-falling-on-regulatory-fears-11621611655.

6 Dirk G. Baur, KiHoon Hong, and Adrian D. Lee. Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions and Money, 54:177–189, 2018. doi:10.1016/j.intfin.2017.12.004.

7 Bruno Biais, Christophe Bisiere, Matthieu Bouvard, Catherine Casamatta, and Albert J. Menkveld. Equilibrium Bitcoin Pricing. Number 3261063. Rochester, NY, 2022. URL: https://papers.ssrn.com/abstract=3261063.

8 Nicola Borri. Conditional tail-risk in cryptocurrency markets. Journal of Empirical Finance, 50:1–19, 2019. doi:10.1016/j.jempfin.2018.11.002.

9 George EP Box and David R Cox. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological), 26(2):211–243, 1964.

10 Brad Chase and Ethan MacBrough. Analysis of the xrp ledger consensus protocol, 2018. URL: https://arxiv.org/abs/1802.07242, doi:10.48550/ARXIV.1802.07242.

11 Nicolas Christin. Traveling the silk road: a measurement analysis of a large anonymous online marketplace. In Proceedings of the 22nd international conference on World Wide Web, WWW '13, page 213–224, New York, NY, USA, 2013. Association for Computing Machinery. doi:10.1145/2488388.2488408.

12 Pavel Ciaian, Miroslava Rajcaniova, and d'Artis Kancs. The economics of bitcoin price formation. Applied Economics, 48(19):1799–1815, 2016. doi:10.1080/00036846.2015.1109038.

13 Lin William Cong, Ye Li, and Neng Wang. Tokenomics: Dynamic adoption and valuation. The Review of Financial Studies, 34(3):1105–1155, 2021. doi:10.1093/rfs/hhaa089.

14 David A. Dickey and Wayne A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica, 49(4):1057–1072, 1981. doi:10.2307/1912517.

15 Paul Kiernan Duehren and Andrew. Bitcoin price surges on biden's crypto executive order. Wall Street Journal, 2022. URL: https://www.wsj.com/articles/biden-to-order-study-of-cryptocurrency-risk-creation-of-u-s-digital-currency-11646823600.

16 Financial Stability Board. Assessment of risks to financial stability from crypto-assets, 2022. URL: https://www.fsb.org/2022/02/assessment-of-risks-to-financial-stability-from-crypto-assets/.

17 Financial Stability Board. Fsb work programme for 2022, 2022. URL: https://www.fsb.org/2022/03/fsb-work-programme-for-2022/.

18 Maddala G.S. Introduction to Econometrics, 3rd Edition. John Wiley & Sons, 2007.

19 James Hamilton. Time Series Analysis. Princeton University Press, Princeton, NJ, 1994.

20 Bruce Hansen. Econometrics. Princeton University Press, Princeton, NJ, 2022.

21 Anna Hirtenstein. Binance crypto exchange ordered to cease u.k. activities. Wall Street Journal, Jun 2021. URL: https://www.wsj.com/articles/binance-crypto-exchange-ordered-to-cease-u-k-activities-11624812672.

22 Tara Iyer. Cryptic Connections. Global Financial Stability Notes. 2022. URL: https://www.imf.org/en/Publications/global-financial-stability-notes/Issues/2022/01/10/Cryptic-Connections-511776.

23 Ryan James. Xrp continues gains following 40% gain on saturday, Apr 2021. URL: https://beincrypto.com/xrp-continues-gains-following-40-gain-on-saturday/.

24 Paraskevi Katsiampa, Shaen Corbet, and Brian Lucey. High frequency volatility co-movements in cryptocurrency markets. 62:35–52, Sep 2019. URL: https://www.sciencedirect.com/science/article/pii/S104244311930023X, doi:10.1016/j.intfin.2019.05.003.

25 Paraskevi Katsiampa, Shaen Corbet, and Brian Lucey. Volatility spillover effects in leading cryptocurrencies: A bekk-mgarch analysis. Finance Research Letters, 29:68–74, Jun 2019. URL: https://www.sciencedirect.com/science/article/pii/S1544612318308237, doi:10.1016/j.frl.2019.03.009.

26 Daisuke Kawai, Alejandro Cuevas, Bryan Routledge, Kyle Soska, Ariel Zetlin-Jones, and Nicolas Christin. Is your digital neighbor a reliable investment advisor? In Proceedings of the ACM Web Conference 2023, WWW '23, page 3581–3591, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3543507.3583502.

27 Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. Scientific Reports, 3(1):3415, 2013. doi:10.1038/srep03415.

28 Yukun Liu and Aleh Tsyvinski. Risks and returns of cryptocurrency. The Review of Financial Studies, 34(6):2689–2727, 2021. doi:10.1093/rfs/hhaa113.

29 James Mackintosh. Behind bitcoin price gyrations: Rational action and wild speculation. Wall Street Journal, May 2021. URL: https://www.wsj.com/articles/bitcoin-is-the-apogee-of-rational-speculation-11621524833.

30 D. L. McLeish. Dependent Central Limit Theorems and Invariance Principles. The Annals of Probability, 2(4):620 – 628, 1974. doi:10.1214/aop/1176996608.

31 Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. IMC '13, page 127–140. Association for Computing Machinery, 2013. doi:10.1145/2504730.2504747.

32 T. Moore and N. Christin. Beware the middleman: Empirical analysis of Bitcoin-exchange risk. In Proceedings of IFCA Financial Cryptography'13, Okinawa, Japan, April 2013.

33 Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008. URL: https://bitcoin.org/bitcoin.pdf.

34 Emiliano Pagnotta. Decentralizing Money: Bitcoin Prices and Blockchain Security. Rochester, NY, 2020. URL: https://papers.ssrn.com/abstract=3264448.

35 Emiliano Pagnotta and Andrea Buraschi. An Equilibrium Valuation of Bitcoin and Decentralized Network Assets. Rochester, NY, 2018. URL: https://papers.ssrn.com/abstract=3142022.

36 M. Hashem Pesaran, Yongcheol Shin, and Richard J. Smith. Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, 16(3):289–326, 2001. doi:10.1002/jae.616.

37 B. Pfaff. Analysis of Integrated and Cointegrated Time Series with R. Springer, New York, second edition, 2008. ISBN 0-387-27960-1. URL: https://www.pfaffikus.de.

38 Peter C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. Biometrika, 75(2):335–346, 1988. doi:10.2307/2336182.

39 Thomas P. Ryan. Modern Regression Methods, chapter 4, pages 146–189. John Wiley & Sons, Ltd, 2008. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470382806.ch4, doi:10.1002/9780470382806.ch4.

40 Linda Schilling and Harald Uhlig. Some simple bitcoin economics. Journal of Monetary Economics, 106:16–26, 2019. doi:10.1016/j.jmoneco.2019.07.002.

41 Gideon Schwarz. Estimating the dimension of a model. The Annals of Statistics, 6(2):461–464, 1978. URL: https://www.jstor.org/stable/2958889.

42 Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In 9th Python in Science Conference, 2010.

43 Michael Sockin and Wei Xiong. A Model of Cryptocurrencies. Number 26816 in Working Paper Series. 2020. URL: http://www.nber.org/papers/w26816, doi:10.3386/w26816.

44 K. Soska, J.-D. Dong, A. Khodaverdian, A. Zetlin-Jones, B. Routledge, and N. Christin. Towards understanding cryptocurrency derivatives: A case study of BitMEX. In Proceedings of the 30th Web Conference (WWW'21), 2021.

45 Andrew Urquhart. What causes the attention of bitcoin? Economics Letters, 166:40–44, 2018. doi:10.1016/j.econlet.2018.02.017.

46 U.S. Commodity and Futures Trading Commission. Cftc charges binance and its founder, changpeng zhao, with willful evasion of federal law and operating an illegal digital asset derivatives exchange | cftc, Mar 2023. URL: From:https://www.cftc.gov/PressRoom/PressReleases/8680-23.

47 U.S. Commodity Futures Trading Commission. Cftc charges sam bankman-fried, ftx trading and alameda with fraud and material misrepresentations, August 2022. URL: https://www.cftc.gov/PressRoom/PressReleases/8638-22.

48 U.S. Commodity Futures Trading Commission. Federal court orders bitmex to pay $100 million for illegally operating a cryptocurrency trading platform and anti-money laundering violations, August 2022. URL: https://www.cftc.gov/PressRoom/PressReleases/8412-21.

49 U.S. Securities and Exchange Commission. Remarks before the aspen security forum, 2021. urlhttps://www.sec.gov/news/public-statement/gensler-aspen-security-forum-2021-08-03.

50 U.S. Securities and Exchange Commission. Sec charges ripple and two executives with conducting $1.3 billion unregistered securities offering, 2021. URL: https://www.sec.gov/news/press-release/2020-338.

51 U.S. Securities and Exchange Commission. Sec files 13 charges against binance entities and founder changpeng zhao, Jun 2023. URL: https://www.sec.gov/news/press-release/2023-101.

52 William W.S. Wei. Time Series Analysis. Oxford University Press, 2013. doi:10.1093/oxfordhb/9780199934898.013.0022.

A Cryptocurrency prices

This section shows the daily average prices and realized volatilities of the cryptocurrencies we consider: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), and Dogecoin(DOGE).

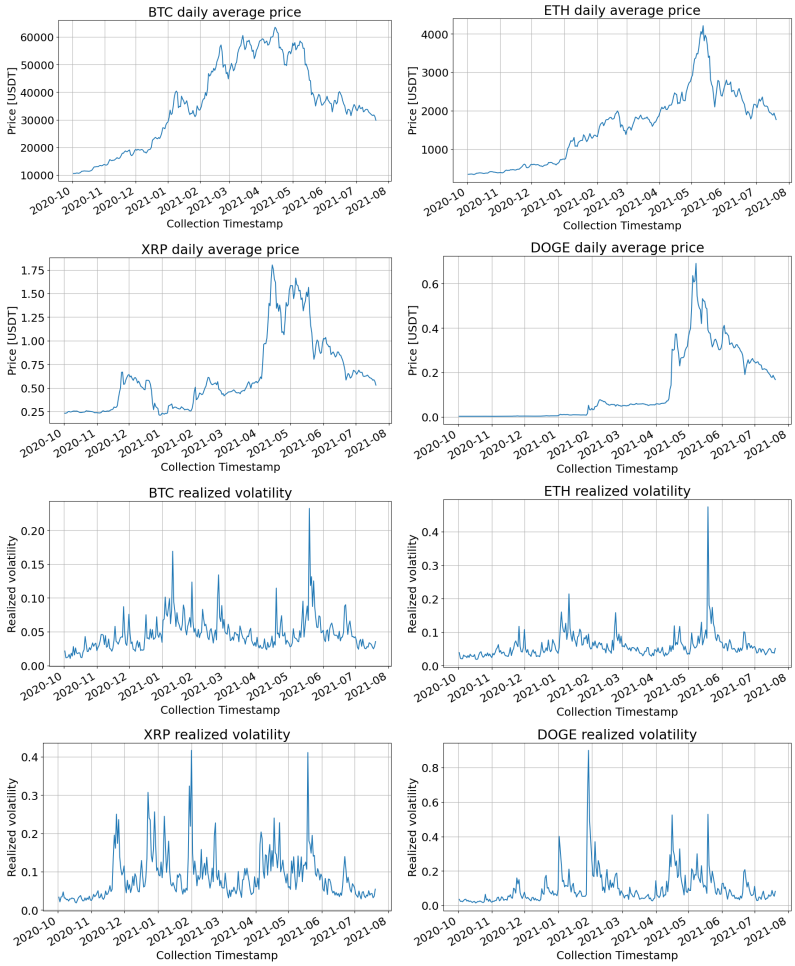


Figure 7 The daily average prices (upper panels) and realized volatilities (lower panels).

B Convergence of marginal effects

This section considers the convergence of marginal effects $\left(\lim_{k\to\infty} \frac{\partial \widehat{\log N_{d+k}}}{\partial V_{X,d}} = 0\right)$.

We can derive the difference equation for a marginal effect from Eqn. (1) and substitute the coefficients with the estimates in Models 1–5 summarized in Table 6. For example, the equation for the first principal component of daily average prices ($v_{1,d}$) in Model 5 is:

$$\frac{\partial \widehat{\log N_{d+k}}}{\partial v_{1,d}} = (1+\pi_0)\frac{\partial \widehat{\log N_{d+k-1}}}{\partial v_{1,d}} = (1-0.712)\frac{\partial \widehat{\log N_{d+k-1}}}{\partial v_{1,d}} \quad (k \geq 2)\,. \tag{5}$$

It clearly shows that the marginal effect converges to zero as $k \to \infty$. We can similarly consider the convergence of every marginal effect and confirm that all marginal effects converge to zero in the limit $k \to \infty$. This means we can consider long-run multipliers for all explanatory variables.