Essays on Cryptocurrency

Inaugural Dissertation to Obtain the Academic Degree of a Doctor in Business Administration at the University of Mannheim

submitted by

Yanghua Shi, M. Sc.
Mannheim, Germany

Dean:	Joachim Lutz
Referent:	Prof. Dr. Erik Theissen
Correferent:	Prof. Dr. Frederik Armknecht

Day of oral examination: June 10, 2025

To My Late Friend

Acknowledgements

In these brief lines, I can only begin to express the depth of my gratitude to the individuals in both my personal and professional spheres whose contributions have played an integral role in making this thesis a reality.

First of all, I would like to thank my supervisor and co-author Prof. Dr. Erik Theissen. Through an excellent balance of giving me plenty of freedom to pursuing my own research interests while at the same time providing continuous guidance, helpful and constructive criticism, and pertinent answers to any of my question, he has tremendously helped me realize this and other projects. Second, I want to thank Prof. Dr. Frederik Armknecht for joining my dissertation committee and providing valuable comments, helping me improve the projects presented in this thesis.

Next, I thank my talented co-authors Dr. Stefan Scharnowski, Fabian Eska, and his supervisor Prof. Dr. Marliese Uhrig-Homburg, for their invaluable contributions to this thesis. I have benefited considerably from their valuable insights and am especially indebted to Stefan for inviting me to collaborate on two projects in cryptocurrency. Many thanks go to my classmates from the Finance Track of the Center of Doctoral Studies in Business for a great time of common studies and course work: Büşra Eroğlu and Kai Mäckle. I also want to thank my current and former colleagues at the Chair of Finance and the Finance Area: Annabelle Bröstl, Büşra Eroğlu, Can Yilanci, Chen Lin, Chia-Yi Yen, Frederik Horn, Dr. Hamed Davari, Hongting Jiang, Justus Veehof, Leah Zimmerer, Lukas Mertes, Mengnan Wu, Minrui Gong, Rongrong Wang, Santanu Kundu, Sven Vahlpahl, Dr. Thomas Johann, and Yufang Sun. Thank you for your excellent cooperation and support in research and teaching. Many thanks also go to our secretary Marion Baierlein and Angelika Wolf-Tobaben.

I am also deeply grateful to the international research community both within and outside of the University of Mannheim. Many thanks go to all the discussants and conference and seminar participants who are too numerous to list here, yet whose feedback has been indispensable to this thesis. I also thank Mannheim's Graduate School of Economic and Social Sciences, especially for the excellent teaching during the first years of my doctoral studies.

Most of all, I want to thank my family: My parents, my grandmother and my aunt. I sincerely appreciate their lifelong support and encouragements, which made this project possible.

Mannheim, December 2023 *Yanghua Shi*

Contents

List of Figures

List of Tables

Introduction

Cryptocurrencies, powered by blockchain technology, have disrupted traditional notions of currency, finance, and transactional systems. Bitcoin, the pioneering cryptocurrency launched in 2009, paved the way for a myriad of alternative cryptocurrencies, each with unique features and use cases. Ethereum, for instance, introduced smart contracts, which enable programmable and self-executing agreements on its blockchain. They have not only introduced a novel avenue for investment that facilitates continuous trading around the clock globally, but they have also forged a unique economic landscape marked by the absence of a central authority controlling the currency. This departure from the conventional norms of the traditional financial system presents both an innovative opportunity and a distinctive challenge for participants in this evolving financial ecosystem. As cryptocurrencies are getting more popular, their impact on global economies and financial institutions becomes more pronounced. This surge, coupled with the apparent absence of fundamental value in cryptocurrencies and their overall high volatility, has ignited discussions among scholars, regulators, and individual participants regarding their viability as an alternative investment. US and European Union have already taken steps towards regulating cryptocurrencies to support the market integrity and financial stability of cryptocurrencies (Browne and Sigalos, 2024; ESMA, 2023). Moreover, the resource-intensive process of mining, used to validate transactions in certain cryptocurrencies such as Bitcoin, which is still the dominant player in the market, has resulted in debates about the sustainability of these cryptocurrencies.

Therefore, this dissertation seeks to contribute to the ongoing discussion about cryptocurrencies by providing a comprehensive examination of their various facets related to investment. In doing so, it bolsters the scientific foundation necessary for not only making informed investment decisions, but also making informed policy decisions and thereby advancing the crucial task of improving regulation. Although cryptocurrency is a relatively new topic, there is already a variety of research streams in this field, as illustrated in this thesis. The first two chapters investigate cryptocurrencies from the point of view of asset pricing, where we attempt to find which design features of cryptocurrencies are relevant for their market capitalization and return volatility. Chapter 3 focuses on one of the design features introduced in Chapter 1 and 2: the consensus mechanism used for transaction validation within the cryptocurrency network. In particular, we investigate the centralization of cryptocurrencies through an event study. This paper is in the field of economic implications arising from blockchain, similar to what was done by Lehar and Parlour (2022). Chapter 4 is in the field of behavioral finance with a more temporal focus, where we investigate intraday investor herding behavior. With many retail investors, herding is particularly relevant for cryptocurrency as an alternative investment and potential store of value. These topics are relevant not only for cryptocurrency investors, but also for regulators and cryptocurrency developers. Table 1 provides an overview of the four chapters by outlining the respective research question, sample periods and data sources, and the key methodologies involved. Subsequent paragraphs give a more detailed summary of each of the chapters.

Table 1: Dissertation Overview

Chapter	Research Question	Period	Data Sources	Methodologies
1	Do design features explain the market capitalization of cryptocurrencies?	(The fourth quarter of) 2020	Kaiko, data from various exchanges, Messari, whitepapers, official network websites, developers' documentation and code repository	Two-stage cross-sectional regression inspired by Karnaukh et al. (2015) LASSO
2	Do design features explain the volatility of cryptocurrencies?	2020 and 2021	Kaiko, data from various exchanges, whitepapers, official network websites, developers' documentation and code repository	LASSO
3	What would be the impact on the Bitcoin network if a blackout were to occur in a region heavily involved in Bitcoin mining?	2021	Exchange data from Kraken Blockchain data	Difference-in-difference
4	Do cryptocurrencies exhibit intraday herding, and if they do, what discernible patterns characterize it?	2017 - 2022	Exchange data from Kraken Google and Reddit data Blockchain data	Newey and West (1987) Estimator OLS

This table provides an overview of the four chapters of this dissertation. *Research Question* broadly summarizes the questions addressed. *Period* describes the sample period. *Data Sources* lists the primary data sources used in the chapter (though additional data sources might be used as indicated in the respective chapters). *Methodologies* lists the main econometric tools used to analyze the data.

Chapter 1 and 2. We note the various design features of cryptocurrencies and their huge cross-sectional variation in the market values and expect that their design features can be relevant to their market performance, analogous to firm fundamentals. In the first two chapters "Design and Valuation of Cryptocurrencies" and "Design and Volatility of Cryptocurrencies", we examine how the design features of cryptocurrencies impact their market capitalization and volatility respectively. We first propose a taxonomy that divides the design features into six groups in the first paper. Then we hand-collect the data of the design features correspondingly. By measuring market performance as market capitalization, we found in the first paper that the cryptocurrencies that are forked from another cryptocurrency and those whose design deviates from the one of Bitcoin tend to have lower market valuation. On the other hand, the non-anonymous cryptocurrencies and those that do not pass on any transaction-specific fees to agents who maintain the integrity of the network, i.e. miners and stakers, have comparably higher market value, due to respectively conformity to Know-Your-Customer and better robustness of the system. In the second paper, we use return volatility to show that older cryptocurrencies and those that are created as the means of payment tend to be less volatile. We also found that cryptocurrency volatility also tends to depend on the type of developer.

Chapter 3. We examine whether Bitcoin is as decentralized as it proclaims using an event study. Miners of proof-of-work networks such as Bitcoin tend to gravitate towards countries with cheap energy. We analyze risks associated with this geographical centralization by exploiting a local electricity supply shock. Compared to a control group consisting of an energy-efficient proof-of-stake cryptocurrency, the Bitcoin blockchain's capacity for transactions decreases while transaction fees increase substantially. The increased settlement latency on the blockchain also reduces secondary market quality as seen in higher exchange rate volatility, lower liquidity, and larger price differences between exchanges. Overall, our results suggest that geographical centralization poses short-lived, but potentially severe system-wide risks to proof-of-work networks.

Chapter 4. We examine the intraday herding of the cryptocurrency market, following the approach by Chang et al. (2000). Due to cryptocurrency markets' relative lack of fundamental information and a large share of retail investors, herding is particularly relevant for cryptocurrency markets. Different from traditional financial markets, cryptocurrency markets are decentralized and open around the clock. This makes the investigation of herding patterns throughout the day possible. We find that herding is strongest during the overlap of hours when traders in major economic centers are likely awake; we also find that herding is stronger when market returns are positive. We also take investor attention and blockchain activity into consideration and find that these two factors have similar patterns as attention and blockchain activity. Additionally, we measure the cross-sectional dispersion of investor attention. To the best of our knowledge, this paper is the first to relate the concept of attention dispersion to the context of investor herding. We find that herding behavior is negatively related to the cross-sectional dispersion of investor attention. Herding behavior is also negatively related to the level of investor attention.

Chapter 1

Design and Valuation of Cryptocurrencies[1]

1.1 Introduction

Cryptocurrency values are highly volatile. While the time-series variation of cryptocurrency values in general (and that of Bitcoin in particular) attracts a lot of public attention, the cross-sectional differences of cryptocurrency values receive much less attention and are not well understood. Some cryptocurrencies, the so-called stablecoins, are backed by a portfolio of assets (and thus have valuations linked to those of the backing portfolios), but most are not. The question therefore arises of what determines the relative valuations of different cryptocurrencies. This question is of obvious importance to users of and investors in cryptocurrencies, to trading venue operators and regulators.

The present paper sheds light on a specific aspect of this issue. We analyze empirically whether design features of cryptocurrencies and the specific characteristics associated with them affect their relative valuation. To this end we first develop a taxonomy of the broad variety of cryptocurrency design features and sort them into six groups, namely, (i) features related to the development process of the cryptocurrency, (ii) technical design features, (iii) features related to cryptocurrency supply, (iv) features related to transactions and transaction processing, and (v) features related to the usability of the underlying network as well as (vi) general features. Additionally, we include the age of each cryptocurrency to take into ac-

[1]This work is the result of collaboration with Prof. Dr. Erik Theissen, Fabian Eska, and his supervisor Prof. Dr. Marliese Uhrig-Homburg. As a group of co-authors, we thank the participants of the 27th Annual Meeting of the German Finance Association (DGF) in Innsbruck, Austria, the Cryptocurrency Research Conference 2022 in Durham, UK, the 11th International Conference of the Financial Engineering and Banking Society in Portsmouth, UK, the FERN seminar of the Department of Economics and Management at Karlsruhe Institute of Technology (KIT), the Ghent Workshop on Fintech 2023 in Ghent, Belgium, and the Economics of Financial Technology Conference 2023 in Edinburgh, UK, for valuable comments and discussions. Further, we thank Deutsche Forschungsgemeinschaft (DFG) for financial support under grants TH 724/7-1 and UH 107/5-1.

count the fact that older cryptocurrencies may have more users and, because of the network externalities associated with the number of users, may be more valuable. We hand-collect a dataset covering the design features and age of 79 cryptocurrencies[2] with the highest market capitalization as of September 2020.

We combine the data on design features with data on market capitalization, obtained by multiplying coin supply by coin prices. To take into account the overall market movements between the "birth" of a coin and our sample period, which is from October to December 2020 in the main analysis, we additionally introduce and analyze a discounted version of market capitalization.

Our dataset is characterized by a high number of potentially relevant independent variables relative to the number of cryptocurrencies in the sample. We use two methodological approaches to tackle this problem. First, we implement a two-stage cross-sectional regression approach inspired by Karnaukh et al. (2015). In step 1 we estimate six regressions in which we regress the market values of the cryptocurrencies in our sample on the design features contained in one of the six groups introduced above. In step 2 we estimate an encompassing regression in which we include those design features that have the highest explanatory power in the respective first-stage regression. Our second approach is the machine learning-based LASSO (least absolute shrinkage and selection operator) regression approach which combines variable selection and regularization. Our approach has two distinct characteristics which differentiate it from traditional asset pricing approaches. First, we explain the cross-section of market valuations, not the cross-section of returns. Second, we do not use a panel dataset (or a repeated cross-section as in Fama and MacBeth (1973)) but rather a simple cross-section. This approach is warranted because our dependent variables (cryptocurrency market values) are highly persistent and most of our independent variables (the design features) have little or no time-series variation.

Our results indicate that cryptocurrencies with a Bitcoin-like combination of design features tend to have higher market capitalization than currencies that are distinctively different from Bitcoin. We also find that cryptocurrency networks that were spun off another network (so-called forks) and not built from scratch tend to have lower market capitalization, possibly because forks compete against their parent networks which are very similar and have a first-mover advantage. Cryptocurrencies that do not pass on any transaction fees and/or tips to agent who maintain the integrity of the network have, on average, a higher market capitalization. Such transaction fees can increase the fragility of the system: some users drop out directly, waiting times increase as a result, and consequently, even more users drop out (Easley et al., 2019; Huberman et al., 2021; Basu et al., 2019). Adverse effects related to network security are also conceivable (Pagnotta, 2022). Our analysis also indicates that networks that require the disclosure of the real-world identities of their users have higher market capitalization. A possible reason for the higher valuation of non-anonymous currencies is that market participants price in the expectation of regulatory approval of non-

[2]We only consider cryptocurrencies in the strict sense, i.e. coins, and exclude tokens because tokens do not operate on their own independent distributed ledger. Even though stablecoins have "coins" in their name, they are generally tokens operating on an existing distributed ledger and therefore are excluded from our analysis.

anonymous currencies and/or regulatory action against anonymous currencies. Finally, we find (weak) evidence that currencies which had for-profit companies as their main developers have lower market capitalization, possibly because of a lower degree of decentralization.

Our paper contributes to the literature on the valuation of cryptocurrencies. A first strand of this literature addresses the question why cryptocurrencies which are neither backed by a pool of assets nor by a trustworthy institution such as a central bank have a non-zero value (Abadi and Brunnermeier, 2018; Aoyagi and Adachi, 2018; Biais et al., 2020; Bolt and van Oordt, 2020; Dwyer, 2015; Pagnotta, 2022; Schilling and Uhlig, 2019; Sockin and Xiong, 2023; Zimmerman, 2020a). A second strand of the literature analyzes financial markets-related determinants of cryptocurrency values. Papers in this area analyze, for example, whether there are common factors driving cryptocurrency returns (Bianchi et al., 2022; Borri et al., 2022; Cai and Zhao, 2021; Hu et al., 2019; Leong and Kwok, 2023; Liu et al., 2020, 2022; Zhang et al., 2021), or whether macroeconomic or regulatory events affect cryptocurrency prices (Auer et al., 2021; Corbet et al., 2020; Koenraadt and Leung, 2022; Li and Miu, 2023). Some papers in this strand of the literature also include cryptocurrency-specific factors driven by network effects or cryptocurrency production cost (Bianchi and Babiak, 2021; Bhambhwani et al., 2019; Cong et al., 2022; Liebi, 2022; Liu and Tsyvinski, 2021). The third strand of the literature, and the one most closely related to our paper, attempts to identify determinants of the cross-section of cryptocurrency values related to cryptocurrency design and blockchain functionality. Two early papers that relate cryptocurrency design to price levels and returns are Hayes (2017) and Wang and Vergne (2017). Hayes (2017) investigates the impact of cryptocurrency design features on prices. He considers prices on a single day in 2014 and four design features, two of which (the rate of coin creation and the use of the scrypt algorithm) turn out to be significant for price formation.[3] Wang and Vergne (2017), in contrast, analyze the returns of five cryptocurrencies and find that they are positively related to a measure of innovation potential as well as to supply growth and liquidity. Further, Shams (2020) shows that the comovement structure of cryptocurrencies is too high to be explained by similarities in characteristics such as the consensus mechanism. He suggests that trading on cryptocurrency exchanges is the main driver of the comovement. We extend this line of research by analyzing a much broader dataset both in terms of cryptocurrencies and in terms of design features, by implementing two distinctively different empirical methodologies and various model specifications, and by proposing a novel taxonomy of cryptocurrency design features.

The remainder of this paper is structured as follows. In Section 1.2, we introduce our novel taxonomy of cryptocurrency design features, describe the data collection procedure, and present descriptive statistics on cryptocurrency design. In Section 1.3, we describe the methodology and in Section 1.4, we present and discuss the results of our empirical analysis. Section 1.5 concludes.

[3] Hayes (2017) also finds that the hashrate affects prices. The hash rate, however, is not a design feature of a cryptocurrency but rather a market outcome.

1.2 Cryptocurrency Design Features

In this section we introduce in Subsection 1.2.1 a novel taxonomy of cryptocurrency design features and hypothesize how the design features might affect the market value of a cryptocurrency. In Subsection 1.2.2 we describe how we collected data on the design features for a total of 79 cryptocurrencies and in Subsection 1.2.3 we present summary statistics.

1.2.1 Taxonomy

The different coins in the cryptocurrency universe can be characterized by combinations of various design features. While there exists a wide range of such features, these can be categorized into a small number of groups. The taxonomy we propose in this section differs from previous attempts which either do not allow a unique allocation of individual features to groups (Garriga et al., 2020), or which create abstract categories difficult to link to individual design features (Cousins et al., 2019). We propose six categories respectively denoted *Development*, *Technical*, *Supply*, *Transactions*, *Usability*, and *General*.

Development
During the development process of a cryptocurrency a basic concept is transformed into implementable code. The identity of the developers and the organization of the process may affect the design and subsequent valuation of the cryptocurrency. With respect to the identity of the developers we differentiate between (i) a loose connection of independent developers and development teams (*DeveloperPublic*), (ii) a non-profit organization (NPO) (*DeveloperNPO*), or (iii) a private, for-profit company (*DeveloperPrivate*). It is not a priori clear how the identity of the development team will affect valuation. On the one hand, users may prefer a public development team because everyone can contribute to the development process, resulting in a high degree of decentralization, a particularly appealing cryptocurrencies characteristic. On the other hand, private developers may have a stronger incentive to make design choices that result in high valuation while public development teams may maximize welfare, which is not necessarily the same thing.

A closely related aspect is the question who decides on code changes. In some networks the decision whether a suggested modification is integrated into the core code is made by the network members and thus by majority voting. We define the dummy variable *MajorityChanges* which is set to 1 for the respective cryptocurrencies.[4] In other networks, a privileged group decides on code changes. For those cryptocurrencies the variable *MajorityChanges* is set to 0. We expect that users value decentralized decision making and that, consequently, cryptocurrencies with majority voting will be more valuable.

We also record general code-related features such as the core code's primary implementation language and the accessibility of the core code. With respect to the implementation language

[4]For *MajorityChanges* to take a value of one, we do not require that *all* decisions on code changes are made by the network members. We only require that some of the decisions are made that way.

we differentiate between C++ (dummy variable *CodeC++*), Go (*CodeGo*), and other languages (*CodeOther*). With respect to accessibility we record whether the core code is fully accessible on Github or a similar platform. If this is the case we set the dummy variable *CodePublicGithub* to 1, and to 0 else. We expect that the lower transparency associated with a non-accessible implementation lowers market capitalization.

The last design feature in the category development is *Fork*, a dummy variable that indicates whether the initial implementation of a cryptocurrency network was forked from another network ($Fork = 1$) or built from scratch ($Fork = 0$). Forks usually improve some aspects of the parent network, possibly resulting in higher valuation of the fork. On the other hand, though, a fork is essentially a (modified) imitation of the parent which lacks innovation and has to overcome the first mover advantage of the parent network. Our prior expectation is that the second effect dominates the first, resulting in lower valuations of forks.

Technical
The category technical comprises design features related to the consensus mechanism, to the hash function, and to the cryptographic methods used to authenticate signatures.

The consensus mechanism provides the rules for reaching agreement on the network status among its users and thus determines how transactions are validated. Validating a transaction is tantamount to authorizing a change to the distributed ledger that documents the change in ownership of the coins transacted. The first and most prominent consensus mechanism is "Proof-of-Work" (PoW), proposed by Nakamoto (2008a).[5] The PoW mechanism results in extremely high energy consumption.[6] Currently, the most important alternative to PoW is "Proof-of-Stake" (PoS). PoS bases on the idea that agents with higher coin holdings are generally more interested in a healthy network. In line with this incentive, the probability that a network member can authorize a transaction is positively related to the coin holdings of that member. Next to (i) PoW and (ii) PoS, Irresberger et al. (2020) identify three further main consensus mechanisms: (iii) Hybrid PoW/PoS, (iv) Delegeted Proof-of-Stake (dPoS), and (v) non-standard consensus mechanisms. We condense these five consensus mechanisms into three dummy variables for PoW (*ConsensusPoW*), for PoS and dPoS (*ConsensusPoSdPoS*), and for nonstandard mechanisms (*ConsensusOther*). We combine PoS and dPoS into one variable because both are based on the aforementioned idea that "richer" network members are more interested in the success of the network and should therefore have more influence on the validation process.[7] We capture hybrids between PoW and PoS by setting both *ConsensusPoW* and *ConsensusPoSdPoS*, to one for the respective cryptocurrencies. The lower energy consumption of PoS is a clear benefit and may result in higher valuation of cryptocur-

[5]In PoW, consensus is reached through the work of so-called miners who compete to solve cryptographic puzzles.

[6]For instance, in 2022 the total electricity consumption of Bitcoin, the most prominent cryptocurrency based on PoW, summed up to about 107.65 TWh according to the Cambridge Bitcoin Electricity Consumption Index (see cbeci.org/) - a value roughly equal to the aggregated electricity consumption of the Netherlands (113 TWh in 2021) according to U.S. Energy Information Administration. Mora et al. (2018) argue that the carbon emissions caused by Bitcoin mining can push global warming above 2°C.

[7]The difference between PoS and dPoS is the fact that in dPoS the network member can outsource the task to third parties, so-called delegates.

rencies adopting that mechanism. However, it is not clear that PoS and other alternative mechanisms are as resistant to attacks as PoW.[8] We therefore have no clear prediction on the sign of the coefficients for the three consensus mechanism dummy variables.

Transactions are combined into blocks, and hash functions are used to ensure that blocks cannot be changed.[9] Within the cryptocurrency universe, many different hash functions are used for this process. For our design feature dataset, we separate between five different specifications: (i) SHA-256, the function which Bitcoin uses, (ii) Ethash or the closely related keccak256 function, (iii) blake, (iv) scrypt,[10] and (v) other hash functions. While we do collect the corresponding data for all cryptocurrencies in our sample, we do not want to inflate the number of independent variables in our empirical analysis. Therefore, we use an additional variable, the age of the hash function, as a proxy for the quality of the hash function. More recently developed hash functions will typically offer a higher level of security.[11] Consequently, we expect a negative influence of hash function age on market valuation.

Cryptocurrencies use Digital Signature Algorithms (DSA) which are based on elliptic curve cryptography in order to authenticate the signatures of the parties in a transaction. We differentiate between three types of elliptic curves, (i) ECDSA, the curve which is used, among others, by the Bitcoin network, (ii) Ed25519, a widely used alternative,[12] and (iii) other curves. While the DSAs are essential for secure coin transfer, not many network users are aware of the specific differences between the elliptic curves. We therefore do not expect a significant impact on market values.

Supply

The process of supplying cryptocurrencies is very different from the process of supplying fiat currency. While the supply of fiat currency depends on the monetary policy of the respective central bank and is therefore subject to discretionary decisions, the supply of cryptocurrencies is predetermined in most networks. Oftentimes, the growth in coin supply is linked to the process of verifying transactions – agents who successfully participate in the verification process are rewarded with newly created coins. In addition, many cryptocurrencies have a supply cap, implying that the maximum number of coins cannot exceed a predetermined threshold. We capture the existence of such a cap by way of a binary variable *MaxSupply* which takes on the value 1 if a cap exists.

In general, cryptocurrencies can (i) have a fixed supply (a feature captured by the dummy variable *FixedSupply*), (ii) be deflationary (*Deflationary*), or (iii) be inflationary. If the supply is fixed, the number of coins in circulation does not vary over time. For deflationary

[8]As a case in point, before the Ethereum network eventually adopted PoS in September 2022, it was stated on the Ethereum homepage that "[PoS] is still in its infancy, and less battle-tested, compared to [PoW]" (see ethereum.org/en/developers/docs/consensus-mechanisms/pos/).

[9]Each block contains the hash value of the previous block. If a block is changed, its hash value will also change and will then deviate from the value written in the next block. This link between blocks makes ex-post changes to a block easily detectable.

[10]For PoW-based crpytocurrencies, Hayes (2017) finds a positive influence of scrypt on prices.

[11]See e.g. Pfautsch et al. (2020) or https://www.streetdirectory.com/etoday/-ejcluw.html.

[12]For instance, Lisk, Monero and Zcash use this elliptic curve for the authentification of signatures. Ed25519 offers a higher level of anonymity compared to ECDSA.

currencies, the number of coins decreases over time as a result of specific "burn mechanisms". Inflationary currencies come in various forms, characterized by different supply growth schemes. Many cryptocurrencies with increasing supply have a reward reduction similar to that of Bitcoin in place. Consequently, the supply curve is increasing and concave over time, and possibly converges to a predetermined threshold. The dummy variable *InflationaryDecreasing* identifies currencies with that feature. Instead of a reward reduction, cryptocurrencies may have constant rewards, resulting in a linear supply function over time (*InflationaryFixed*). Finally, the supply curve may be convex. This can be achieved by fixing the supply growth rate (rather than the number of coins issued per unit of time). The growth rate is often referred to as the rate of inflation of the currency (*InflationaryFixed-InflationRate*). Finally, some currencies have dynamic and thus time-varying supply growth rates, resulting in non-deterministic supply growth (*InflationaryDynamic*). We expect that cryptocurrencies with supply caps, fixed supply and deflationary currencies have lower value because the supply restrictions may limit the adoption of the currency by users.

As noted, the reward to those agents verifying transactions in the network is linked to coin supply. The reward can be a coinbase reward (the creator of each new entry to the ledger earns a specific number of new coins) or an alternative reward distributions scheme, e.g. one where rewards are distributed among a larger network user group (e.g. all verifiers) and are not necessarily attached to individual new ledger entries. We capture these two cases by the dummy variables *RewardCoinbase* and *RewardAlternative*. In either case, those two reward distribution schemes incentivize agents to contribute to a healthy network and thus, we expect a positive influence on market capitalization.

Transactions

The category transactions contains design features related to transactions on the cryptocurrency network and the ways in which these transactions are processed.

The number of transactions a network can process per period of time is often referred to as the throughput. In theory it can be measured by transactions per second (TPS). However, TPS is controversial, most importantly because it cannot be measured consistently for all networks. Therefore, we proxy TPS by the time between blocks and the existence or non-existence of a blocksize limit. The time between blocks determines the frequency of changes of the distributed ledger. We differentiate between the theoretically intended minimum time between two blocks (variable *TheoreticalBlockTime*) and the actually observed time between blocks (*BlockTimeAverage*).[13] We note that lower time between blocks does not only mean that more transactions can be processed per unit of time, but also means that the minimum time it takes to complete a transaction is lower. We therefore expect that lower time between blocks is associated with higher valuation. A blocksize limit sets a limit to the number of transactions that can be processed per unit of time and thus limits the throughput of the network. The dummy variable *BlocksizeLimit* is set to one if such a limit exists.[14] We expect

[13]If there was a fork within a network that induced a change in at least one of these variables, we record the post-fork values of the variables. In the subsequent analysis, we restrict ourselves to the actually observed blocktime due to data availability and reliability.

[14]We were unable to verify whether a blocksize limit exists for some cryptocurrencies, implying that we have missing data for this variable.

that the existence of a blocksize limit affects market value negatively.[15]

In many cryptocurrency networks users have to pay a fee for the processing of their transactions. We include three variables that intend to capture the existence and design of such fees, *TransactionFeeObligation*, *TipSpecialTreatment*, and *NoFeeTipForMinerForger*. *TransactionFeeObligation* records whether a cryptocurrency network has a mandatory fee for a transaction to be processed. Because the existence of a mandatory fee makes it more expensive to use the network, we expect a negative impact on market valuation. Some networks allow their users to prioritize a transaction by paying a special fee, often called tip. We define the dummy variable *TipSpecialTreatment* which is set to one if tips are possible. We expect that investors value the possibility to prioritize their transactions and therefore expect a positive impact on market valuation. The third variable, *NoFeeTipForMinerForger*, is set to one for networks where the transaction fees and/or the tips are not - neither fully nor partly - passed on to the agents verifying transactions (e.g. miners in PoW and stakers in PoS). A scheme where fees and/or tips are passed on to those agents (miners or stakers) makes their activities more profitable and may thus attract more agents. This, in turn, increases the degree of network decentralization and the security (i.e., resistance against attacks) of the network. We therefore expect a negative effect of *NoFeeTipForMinerForger* on market values.

Usability
The first cryptocurrency, Bitcoin, was devised as a means of payment. However, there are use cases for cryptocurrencies beyond that. A cryptocurrency network can be a payment system, a platform for smart contracts (the Ethereum network is a case in point), or it can serve other purposes such as decentralized finance applications. We capture the intended use of a cryptocurrency by three dummy variables, *IntentionPayment*, *IntentionSmartContract*, and *IntentionOther*. We expect that cryptocurrencies that serve purposes beyond being a means of payment have higher market values.

In some networks the ownership of coins embodies rights (e.g. voting rights), or possibilities of usage beyond making payments. The variable *UsageBeyondPayment* takes on the value one for cryptocurrencies for which this is the case. We expect a positive coefficient. Some cryptocurrency networks offer implicit smart contract support (without requiring sidechains or similar arrangements).[16] For networks with this feature we set the dummy variable (*SmartContractSupport*) to one. We anticipate a positive value impact due to expanded functionalities, but the risk of hacking attacks on smart contracts resulting from implementation errors is likely to introduce a negative effect. The dominant effect remains undetermined.

General
The final category comprises three further design characteristics that may potentially affect valuation. Most cryptocurrencies bundle transactions into blocks and update the network status by appending blocks to a blockchain. Starting with Ripple in 2012, the cryptocur-

[15] We note, though, that an unlimited block size may result in excessively large ledger entries.

[16] There certainly are other features that extend the usability of a cryptocurrency. However, we are not aware of other features that are consistently documented in the public domain. We therefore restrict ourselves to the variable *SmartContractSupport*.

rency universe was extended by networks that do not apply the blockchain technology but an alternative distributed open source protocol. The variable *LedgerOther* identifies such networks. Generally, these alternative designs aim at overcoming the scalability problem of blockchains, thereby potentially creating possibilities for new usages of cryptocurrencies. This aspect might lead to a higher valuation of the respective cryptocurrencies. However, these no-blockchain designs may be less secure (or less battle-proof), limiting their adoption and lowering their valuation. As it is unclear which of these effects is stronger, we do not have a clear prediction for the sign of the coefficient on the *LedgerOther* variable.

There are two accounting schemes that are commonly applied in the cryptocurrency world. The first cryptocurrencies (including Bitcoin) relied on unspent transaction outputs (UTXOs) to balance the ledger. Under this accounting scheme the ledger does not store information on account balances. Consequently, to infer account balances one has to process the entire blockchain and sum up all UTXOs logged to the respective account. Given the enormous size of many blockchains this may not be the most efficient solution. Therefore, other cryptocurrency networks apply a traditional balance accounting scheme. Such networks store every account's balance on the blockchain (similar to banks that store customer account balances using electronic records). This accounting scheme does not require a network member to parse the whole ledger to infer account balances. Rather, a synchronization without accessing the whole history of the ledger becomes possible. We identify cryptocurrencies using such an accounting scheme by the variable *AccountingBalance*. We expect a positive coefficient because of the efficiency and intuitive appeal of these accounting schemes.

An important feature is the degree of anonymity that a cryptocurrency network offers its users. In networks such as Bitcoin, any transaction and amount in a wallet can be traced back to a pseudonym (known as the public address). Other networks focus on the provision of a greater degree of privacy and enable completely anonymous transactions using specific cryptographic methods.[17] We identify networks that allow anonymous transactions by the variable *Anonymous*. The increased level of privacy satisfies a demand for censorship resistance[18] and thus makes anonymous cryptocurrency networks more attractive. We therefore expect that networks supporting full anonymity have higher valuation than those which only allow pseudonymous transactions. On the other end of the anonymity spectrum are cryptocurrency networks that connect the addresses and transactions to real world identities (identified by the variable *NonAnonymous*). Such a non-anonymous design may offer advantages with respect to regulatory acceptance because KYC (Know-Your-Customer) is

[17]Examples include zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Argument of Knowledge) of Zcash, a form of zero-knowledge cryptography. In this network, transactions can be fully encrypted but the validity can still be verified with specific zk-SNARK proofs. In detail, a "prover" can prove to a "verifier" that a statement is true without revealing any information beyond the validity itself. Via specific combinations, this procedure allows transaction processing without disclosing information about the transaction itself.

[18]Pagnotta and Buraschi (2018) state that censorship resistance "has multiple sources including financial repression through governmental capital controls; option-like hedging against government abuses such as arbitrary wealth confiscations or the targeting of political dissidents and/or religious groups; hedging against changes in inheritance laws; the risk of disruptions of the traditional banking system due to bank runs, fiat hyperinflation or forced maturity conversion of bank deposits; the ability to secure wealth transfers in the event of armed conflicts, territorial invasions, civil wars, refugee crises", as well as criminal activity.

already fulfilled. Supervisory authorities might thus favor non-anonymous cryptocurrencies. In addition, note that these networks do not necessarily bear a higher risk to reveal their members' identity to the public. These currencies may have higher values than those with pseudonymous transactions due to the prospect for regulatory acceptance.

In Table 1.1 we provide an overview of the variables we have introduced above. We indicate whether a variable is binary or not, and we document our expectation on the sign of the coefficient ("+" or "-"). A "0" indicates that we either have no clear prediction for the respective variable, or that we consider its impact on valuation to be negligible.

Table 1.1: Design features variables, expected influence, and descriptive statistics

	Binary	Predicted influence	As of December 2020			All-time		
			Obs.	Mean	Std. Dev.	Obs.	Mean.	Std. Dev.
Panel A: *Development*								
DeveloperPublic	yes	+	79	0.2025	0.4045	114	0.2368	0.4270
DeveloperNPO	yes	+	79	0.2785	0.4511	114	0.2544	0.4374
DeveloperPrivate	yes	–	79	0.5190	0.5028	114	0.5088	0.5021
MajorityChanges	yes	+	79	0.6456	0.4814	114	0.6842	0.4669
CodePublicGithub	yes	+	79	0.9620	0.1924	114	0.9737	0.1608
CodeC++	yes	0	79	0.3924	0.4914	114	0.4386	0.4984
CodeGo	yes	0	79	0.3671	0.4851	114	0.3246	0.4703
CodeOther	yes	0	79	0.2532	0.4375	114	0.2456	0.4324
Fork	yes	–	79	0.5063	0.5032	114	0.5526	0.4994
Panel B: *Technical*								
ConsensusPoW	yes	0	79	0.3165	0.4681	114	0.4386	0.4984
ConsensusPoSdPoS	yes	0	79	0.4937	0.5032	114	0.4035	0.4928
ConsensusOther	yes	0	79	0.2278	0.4221	114	0.2018	0.4031
HashSHA256	yes	N/A	79	0.4304	0.4983	114	0.4035	0.4928
HashEthash	yes	N/A	79	0.1519	0.3612	114	0.1316	0.3395
HashScrypt	yes	N/A	79	0.0759	0.2666	114	0.0789	0.2708
HashBlake	yes	N/A	79	0.1392	0.3484	114	0.1140	0.3193
HashOther	yes	N/A	79	0.2785	0.4511	114	0.3421	0.4765
HashAge	no	–	79	4752.99	1993.83	114	4614.67	1994.05
CurveECDSA	yes	0	79	0.6329	0.4851	114	0.6316	0.4845
CurveED25519	yes	0	79	0.3418	0.4773	114	0.3158	0.4669
CurveOther	yes	0	79	0.0759	0.2666	114	0.0877	0.2841
Panel C: *Supply*								
MaxSupply	yes	–	79	0.6582	0.4773	114	0.7105	0.4555
FixedSupply	yes	–	79	0.2278	0.4221	114	0.2105	0.4095
Deflationary	yes	–	79	0.1139	0.3197	114	0.0789	0.2708
InflationaryDecreasing	yes	0	79	0.4177	0.4963	114	0.4825	0.5019
InflationaryFixed	yes	0	79	0.1013	0.3036	114	0.1053	0.3082
InflationaryFixedInflationRate	yes	0	79	0.0506	0.2206	114	0.0439	0.2057
InflationaryDynamic	yes	0	79	0.1772	0.3843	114	0.1404	0.3488
Inflationary	yes	0	79	0.7468	0.4376	114	0.7719	0.4214
RewardCoinbase	yes	+	79	0.6582	0.4773	114	0.6930	0.4633
RewardAlternative	yes	+	69	0.3165	0.4681	114	0.2632	0.4423

	Binary	Predicted influence	As of December 2020			All-time		
			Obs.	Mean	Std. Dev.	Obs.	Mean.	Std. Dev.
Panel D: *Transactions*								
TheoreticalBlockTime (seconds)	no	−	73	99.65	175.53	105	127.16	197.43
BlockTimeAverage (seconds)	no	−	76	97.83	170.27	106	127.55	196.0435
BlocksizeLimit	yes	−	60	0.7333	0.4459	91	0.7473	0.4370
TransactionFeeObligation	yes	−	77	0.7143	0.4547	111	0.6577	0.4766
TipSpecialTreatment	yes	+	73	0.5616	0.4996	105	0.6	0.4922
NoFeeTipForMinerForger	yes	−	79	0.2025	0.4045	114	0.1667	0.3743
Panel E: *Usability*								
IntentionPayment	yes	0	79	0.3291	0.4729	114	0.4035	0.4928
IntentionSmartContract	yes	+	79	0.3671	0.4851	114	0.3070	0.4633
IntentionOther	yes	+	79	0.3038	0.4628	114	0.2895	0.4555
SmartContractSupport	yes	0	79	0.6835	0.4681	114	0.5789	0.4959
UsageBeyondPayment	yes	+	79	0.4430	0.4999	114	0.3947	0.4910
Panel F: *General*								
LedgerOther	yes	0	79	0.0633	0.2450	114	0.0702	0.2566
AccountingBalance	yes	+	79	0.5316	0.5022	114	0.4561	0.5003
Anonymous	yes	+	78	0.2692	0.4464	113	0.2832	0.4526
Pseudoanonymous	yes	−	78	0.7051	0.4589	113	0.6991	0.4607
NonAnonymous	yes	+	78	0.0641	0.2465	113	0.0442	0.2066

Maintaining the different design feature groups, this table lists the variables introduced in Section 1.2.1 and summarizes its expected influence on market capitalization. Additionally, the columns on the right provide descriptive statistics for each variable. We consider (i) all cryptocurrencies in our sample in their design configuration as of December 2020 and (ii) all cryptocurrencies in our sample including all of their historical design feature combinations. Since we do not include the specific hash functions into our empirical analysis, we do not attempt to predict their influences on market valuation.

1.2.2 Data Collection

Unlike price and quotation data for cryptocurrencies, data on design features cannot be obtained from data vendors. We therefore had to hand-collect data on the variables introduced in Section 1.2.1. We used with priority data sources directly related to the network founders, the development team, and the network community. These data sources include, for example, whitepapers, official network websites, developers' documentation, and the code repository. When required information was not available from these sources, or when the information provided was incomplete or inconsistent, we extended our search to expert forums like the respective subreddits and those on the developers' portals.[19] For the variables *IntentionPayment*, *IntentionSmartContract* and *IntentionOther*, we restrict our data collection procedure to the tags provided by Coinmarketcap and Messari. For well-known and highly capitalized cryptocurrencies we could collect the required data rather easily. However, the quality of documentation is often poor for less well known and less capitalized cryptocurrencies. For many of these currencies the data required for our analysis was unavailable in spite of the broad range of data sources we tapped. Eventually we managed to collect data on the relevant design features for the 79 cryptocurrencies with the highest market capitalization as of September 2020. We admit that our dataset is not free from survivorship bias. However, because data on design features of cryptocurrencies with low market valuations and of cryptocurrencies that were discontinued is unavailable, there is no straightforward way to resolve this issue. We try to mitigate it by relating market valuation to *lagged* data on design features, i.e. the data on design features from September 2020, the month right before our sample period.

If we include in our dataset soft forks that imply a change in at least one design features of our taxonomy,[20] our dataset increases from 79 to 114 observations. Note, though, that this all-time dataset includes those cryptocurrencies that experienced a design change that was not associated with a hard fork twice, respectively over different periods, since at each point in time only one of the two versions of the same cryptocurrency existed. We retain (and make available to other researchers) the all-time dataset because it allows to reconstruct the exact design configuration of all cryptocurrencies in the sample at any point in time during the sample period.

Despite our attempts to collect complete data for all cryptocurrencies, there are some variables with missing observations. These include *RewardAlternative* (10 missing observations), *TheoreticalBlockTime* (6 missing observations), *BlockTimeAverage* (3 missing observations), *BlocksizeLimit* (19 missing observations), *TransactionFeeObligation* (2 missing observations),

[19] Whenever we rely on these data sources we require that the information in our dataset is backed by two independent sources.

[20] For instance, Monero, a cryptocurrency which allows completely anonymous transactions by obscuring transaction senders and recipients through cryptography, originally had a blocktime of one minute. In 2016, the blocktime was raised to two minutes (alongside with some other changes not relevant in the context of our design feature variables). Such a situation implies two entries in our all-time dataset. The first one includes the initial blocktime of one minute while the blocktime variable is set to two minutes in the second entry. Such changes are not necessarily associated with hard forks that result in two cryptocurrencies existing simultaneously after the fork date.

TipSpecialTreatment (6 missing observations), and the degree of anonymity (1 missing observation).

1.2.3 Summary Statistics

Table 1.1 shows summary statistics (number of observations, mean and standard deviation) for all design feature variables, both for the all-time dataset (the one that contains soft-forked cryptocurrencies) and for our main dataset containing 79 cryptocurrencies in their design configuration as of December 2020. We describe summary statistics for the latter dataset. This description does not only characterize our sample but "in passing" also provides an overview of the designs of the most important cryptocurrencies. We note that in some cases the categories we have created to capture alternative specifications of a design feature are not mutually exclusive. As a consequence, the fractions shown in Table 1.1 can add up to more than 100%.[21]

We find that approximately half of the cryptocurrencies were developed by private, for-profit entities, 27.9% by not-for-profit organizations, and 20.3% by networks of independent developers. In 64.6% of the cryptocurrency networks decisions on major code changes and/or decisions on governance issues are passed on to the network members. This figure implies that some networks which were developed by for-profit entities still involve the users in the development process. We further find that nearly all networks have publicly available core codes, and that most cryptocurrency networks are either using C++ (∼39.2%) or Go (∼36.7%). About 50% of the cryptocurrencies were forked from another network, while the others were built from scratch.

31.7% of the cryptocurrencies in our sample use a consensus mechanism based on proof of work. Proof of stake or delegated proof of stake are more widely used (49.4%), and 22.8% of the cryptocurrencies use other consensus mechanisms. These figures are in line with the observation, made by Irresberger et al. (2020), that proof of stake is becoming more popular. The most widely used hash function is Bitcoin's SHA-256 (43.0%), followed by HashEthash (15.2%). Although different elliptic curves can theoretically be used for signature generation, most coins use the two standard digital signing algorithms ECDSA (63.3%) and Ed25519 (34.2%).

Of the 79 cryptocurrencies in our sample, 52 (65.8%) have a supply cap. 22.8% of the coins have a fixed supply while 11.4% are deflationary. Of the cryptocurrencies with increasing supply (74.7%), most have adopted a scheme with decreasing growth rates (41.8% of the total, equivalent to 56% of the inflationary currencies). The alternative growth schemes are less popular. In about two thirds of the networks in our sample, the verifying agents are rewarded with coinbase rewards. An alternative reward scheme is used by 31.7% of the

[21] Consider, for example, the three dummy variables which capture the consensus mechanism (*ConsensusPoW, ConsensusPoSdPoS, and ConsensusOther*). Three cryptocurrencies use a combination of consensus mechanisms, for each of which we assign a value of 1 to two of the corresponding variables. Therefore, the means shown in the table add up to 1.038.

networks (note that there are some missing values for this variable).

The summary statistics of the design features in the category "transactions" indicate that the average theoretical blocktime amounts to 99.65 seconds with a standard deviation of 175.53. The blocktime that is actually observed in the market is slightly lower, at 97.83 seconds with a standard deviation of 170.27.[22] Of the 60 cryptocurrencies for which we could infer whether a blocksize limit exists, approximately three quarters (73.3%) have such a limit in place. 71.4% of the networks require their users to pay a mandatory transaction fee, and 56.2% allow a prioritization of transactions by paying a tip. In 20.3% of the cryptocurrency networks in our sample, transaction fees and/or tips are not included in the rewards for miners and stakers.

Turning to the variables in the "usability" group, we find that the original intention of the cryptocurrencies in our dataset is roughly evenly distributed across the categories payment system, smart contract platform, and other. 68.4% of the cryptocurrency networks support smart contracts within their core code implementation, and in 44.3% of the networks coin holdings are associated with further rights, such as voting rights, or enable usages beyond pure transaction purposes.[23]

The overwhelming majority (93.7%) of the networks in our sample use a blockchain-based ledger. Only 6.3% use alternative ledgers. More than half (53.2%) of the cryptocurrencies use a balance accounting scheme, leaving 46.8% for UTXO accounting. With respect to the degree of anonymity, a strong majority (70.5%) of the networks allows pseudonymous transactions (as is also the case in the Bitcoin network). 26.9% are fully anonymous while only 6.4% of the networks require the disclosure of the real world identities of their users.

1.3 Empirical Methodology

1.3.1 Two-Stage Regressions and LASSO

We aim at empirically analyzing which design features from our taxonomy have explanatory power for the cross-sectional valuation of cryptocurrencies. To identify those that significantly affect the value of cryptocurrencies, we regress two different measures of market valuation on the design feature variables introduced in Section 1.2. Our empirical design is characterized by a low number of observations (the 79 cryptocurrencies) and a large number of explanatory variables, such that a standard regression analysis is unlikely to yield reliable results. To tackle the issue of overfiting we use two methodological approaches, a two-step regression approach inspired by Karnaukh et al. (2015) and LASSO (least absolute shrinkage

[22] Reducing the sample "as of December 2020" to the observations for which both, theoretical and actual, blocktimes are available, we have average blocktimes of 99.11 seconds and 99.43 seconds for the theoretical and the observed blocktimes, respectively.

[23] Binance coin is an example of a coin that provides such additional usage. The coins in this network can be used to pay for several fees when using the centralized exchange Binance, such as listing fees.

and selection operator) regressions.

The two-step regression approach proceeds as follows. We first estimate six separate regressions with the respective market valuation measures as the dependent variable and the design features of one of our six categories as independent variables ("intra-group regressions"). Those variables with the highest explanatory power are then included as independent variables in the second-pass regression ("encompassing regression"). We judge the explanatory power by the p-values in the intra-group regressions and use different cut-off values (0.3, 0.2 and 0.1). We further include the age of the cryptocurrencies in the encompassing regression.

The LASSO regressions connect variable selection and regularization.[24] Specifically, we perform a 10-fold cross validation with random subsets selection to determine the tuning parameter λ that minimizes the mean squared error (MSE) for the LASSO regression with intercept. Based on the value of λ, the LASSO is then applied on the entire dataset to determine the model's parameter estimates and the intercept. We repeat this procedure $10,000$ times in order to base our inference on a broad range of different training and validation data subset compositions.[25]

The majority of our independent variables (i.e., the design characteristics) are time-invariant. We therefore use time-series averages of the dependent variables to eliminate effects that may be specific to individual days. As noted in Section 1.2.2, the 79 cryptocurrencies in our sample are those with the highest market capitalization as of September 2020. To alleviate endogeneity issues, we use in our main analysis market valuation data averaged over all days of the fourth quarter of 2020. We show results for alternative specifications in Section 1.4.2.

1.3.2 Variables Definitions

Dependent Variables: Market Valuation Data

We measure market valuation by market capitalization and a discounted version of the market capitalization. The calculation of market capitalization requires data on cryptocurrency prices and circulating supply. We obtain the former from the APIs of the respective exchange and from cryptocurrency data provider Kaiko, while the latter is obtained from Messari.[26]

[24]Compared to a standard linear regression (with intercept) a penalty term $\lambda \sum_{j=1}^{n} \mid \beta_j \mid$ is introduced and the algorithm's objective is to minimize $\sum_{i=1}^{n} \left(y_i - \alpha - \sum_j x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{n} \mid \beta_j \mid$, see e.g. Tibshirani (1996). The choice of λ is crucial. The higher λ, the more variables are eliminated, but the deviation of estimated values and observed data introduced by the approach increases. If λ is low instead, more variables are selected and the variation in the predictions increases.

[25]When we use five folds instead of ten in the cross validation procedure, our results remain qualitatively similar.

[26]Pricing data from Kaiko is used only in case there are missing values in the API data. Messari provides cryptocurrency data and is recommended by Kaiko as a source for circulating supply. Circulating supply excludes coins/tokens from the outstanding supply that are (i) restricted by any contracts, e.g. on-chain-lockups, or (ii) are held by projects/foundations without selling intention (see https://messari.io/report/messari-proprietary-methods). This mitigates concerns that our market capitalization variable actually measures trading motives and/or investor sentiment rather than "equity value".

The price data that we use is a daily volume-weighted average of the prices of nine cryptocurrency exchanges.[27] Whenever a cryptocurrency is traded against USD on an exchange, our dataset which is combined from the exchange API and Kaiko provides daily volume-weighted average prices for the respective venue. We use these prices whenever available, and we refer to them as direct prices. Not every cryptocurrency is traded against USD on every exchange. Therefore, direct prices are not always available. However, these cryptocurrencies are usually traded against BTC, and BTC is traded against USD. We use these two prices to calculate an implicit USD price of the cryptocurrency under consideration and refer to these implicit prices as indirect prices.[28] When doing so we implicitly assume that USD and BTC quotes are consistent. One exchange, Binance, is an exception. It does not trade cryptocurrencies against USD, but it does trade them against EUR. We calculate indirect prices for those cryptocurrencies traded against EUR on Binance by combining the EUR price of the currency with the EUR-USD exchange rate (obtained from exchangerate.host).[29] To check the reliability of the indirect prices we calculate indirect prices for cryptocurrency-exchange combinations for which direct prices are also available. We observe average differences below 1% for almost all combinations.

There are different ways how we could construct our final dataset. We could use direct prices where available and use indirect prices only when direct prices are unavailable, or we could generally use indirect prices. We opt for a combination of these procedures. We use indirect prices when only these are available, and we use a volume-weighted average of direct and indirect prices when both are available.[30] We end up with one price for each cryptocurrency-exchange pair that is either an indirect price or a weighted average of direct and indirect prices. These prices are then used to calculate the weighted average price across the nine exchanges. We then multiply this weighted average price by the circulating supply to obtain our measure of market capitalization (referred to as *plain* market capitalization in the sequel).

The cryptocurrencies in our sample are of very different age. For instance, the genesis block of Bitcoin was created in January 2009 while Avalanche was just introduced in mid-September

[27] We originally obtained data of the following ten exchanges from Kaiko: Binance, Bitfinex, Kraken, Bitstamp, Coinbase, bitFlyer, Gemini, itBit, Bittrex, and Poloniex. These exchanges are considered reliable, meaning that they do not report inflated volume (see Härdle et al. (2020); on the importance of reliable data in the context of cryptocurrency trading data, see Alexander and Dakos (2020)). We exclude Poloniex because the only fiat currency traded on this exchange is Malaysian ringgit (RM). Due to the low liquidity between RM and USD, we refrained from converting the RM quote to a USD quote via the RM-USD rate.

[28] For example, the price of ABBC Coin (ABBC) in USD is not available on Bittrex, but Bittrex trades ABBC against BTC and BTC against USD. With $\frac{USD}{ABBC} = \frac{BTC}{ABBC} \cdot \frac{USD}{BTC}$, we obtain the Bittrex USD price of ABBC coins.

[29] We are aware of the fact that there are arbitrage opportunities in the cryptocurrency market (see e.g. Makarov and Schoar (2020)). We note, though, that the exchanges in our sample are among the most liquid cryptocurrency exchanges, and higher liquidity is usually associated with higher market efficiency. Furthermore, the cryptocurrency market has generally become more efficient over time (Noda, 2021; Köchling et al., 2019; Kristoufek and Vosvrda, 2019).

[30] We do so to alleviate endogeneity concerns. Indirect prices may be systematically biased, and it is more likely that a cryptocurrency with low market capitalization is not directly traded against USD. Note that our results are qualitatively similar when we use indirect prices throughout.

2020. On average, cryptocurrencies that are older and more established are associated with higher market capitalization, possibly because of network effects (Metcalfe, 2013; Alabi, 2017) and/or because older cryptocurrencies tend to be less volatile (Kim, 2015; Hafner, 2020; Nabilou and Prüm, 2019) and thus are better suited to act as a store of value. An older cryptocurrency may have established a "brand value" and customer loyalty. These cryptocurrency networks may also be less impacted by adverse news,[31] possibly resulting in higher market capitalization. Finally, these cryptocurrencies also benefit from the overall enhancement of the whole cryptocurrency market. Therefore, in addition to the "plain" market capitalization we also analyze discounted market valuation. Specifically, we adapt the fund size scaling procedure of Pástor et al. (2015) and calculate the discounted market capitalization of cryptocurrency i at time t according to

$$DiscountedMCap_{i,t} = MCap_{i,t} \cdot \frac{CRIX_{Genesis_i}}{CRIX_t} \tag{1.1}$$

with $CRIX_{Genesis_i}$ and $CRIX_t$ denoting the value of the CRIX (see e.g. Trimborn and Härdle (2018)), a widely used cryptocurrency market index, at the genesis date of coin i and at time t, respectively. For the seven cryptocurrencies in our dataset that were launched prior to the CRIX, i.e. before July 31, 2014, we set $CRIX_{Genesis_i}$ to the CRIX's initial value of 1000. Intuitively, the procedure described by equation (1.1) deflates the value of cryptocurrency i at time t to its launch date.

For either methodological approach, we control for outliers by winsorizing the top three cryptocurrencies according to market capitalization (Bitcoin, Ethereum, and Ripple) and discounted market capitalization,[32] respectively. We rescale the market capitalization variables to the range $[0, 1]$ in order to obtain coefficient estimates of a convenient magnitude. As a robustness check we also estimate an alternative specification that avoids winsorizing. Specifically, we use the log of the plain and discounted market valuation as dependent variables and obtain results (not tabulated) that are qualitatively similar to those reported in the paper.

Independent Variables: Design Feature Data

Irrespective of the methodology applied we furthermore reduce the number of independent variables by conflating some of them. Specifically, we do not include the variables *CodeGo* and *CodeOther* but rather the binary variable *CodeNonC++* which is set to one if at least one of the former variables is one, and zero otherwise. Similarly, we introduce the binary variable *CodeNonECDSA* to identify networks which do not use ECDSA for signature generation. Within the design feature group usability, we combine *IntentionSmartContract* and

[31] Bianchi (2020), Finck (2018), Jo et al. (2020), and Polasik et al. (2015) argue that the cryptocurrency market is heavily sentiment-dependent and its users consider blockchain as an immature technology that is still evolving and whose practical impact is uncertain. As a result, well-performing cryptocurrencies are likely to already exist for a longer time and be less impacted by adverse news.

[32] When considering different time horizons to calculate the time-series average of the discounted market capitalization, we notice that the top three cryptocurrencies are more than five interquartile ranges above the third quartile and therefore should be considered as outliers. Note that the top three cryptocurrencies are not always the same - for example, in the fourth quarter of 2020, the top three cryptocurrencies according to discounted market capitalization are Bitcoin, Polkadot and EOS.

IntentionOther to the new variable *IntentionNonPayment*. Moreover, we do not include the variables that identify the different inflationary supply curves but restrict ourselves to the aggregated variable *Inflationary*. Finally, as already mentioned in Section 1.2.1, we do not include the specific hash function variables into our regression analysis but rather only include the age of the hash function.

Several of our independent variables are exhaustive sets of dummy variables (such as *CodeC++* and *CodeNonC++*). We therefore need to define a base case and exclude the corresponding dummy variable from the regression. We always dropped the variable that corresponds to the design of the Bitcoin network. As an implication of this specification, the constant in our regression captures the value of a network that has a Bitcoin-like combination of those design features captured by the dummy variables, with all other variables equal to zero. We then take this approach one step further and recalculate actual blocktimes as well as the age of the hash function according to

$$BlockTime_{mod} = \frac{TheoreticalBlocktime_{Bitcoin} - Blocktime}{TheoreticalBlocktime_{Bitcoin}} = \frac{600 - Blocktime\,[s]}{600} \tag{1.2}$$

and

$$HashAge_{mod} = \frac{HashAge_{Bitcoin} - HashAge}{HashAge_{Bitcoin}}, \tag{1.3}$$

respectively. These modified variables take on a value close to zero (equal to zero) for the Bitcoin network. Positive values indicate a blocktime lower than that of the theoretical one of Bitcoin (implying higher throughput of the network as compared to Bitcoin), and a hash function that is younger (and thus arguably more secure) than that used by the Bitcoin network, respectively. We note that Bitcoin has the highest theoretical blocktime (10 minutes) and the oldest hash function in our sample (SHA-256), implying that the two modified variables take on the value close to zero for Bitcoin and positive values for any cryptocurrency in our sample. We further rescale the two variables to the interval [0,1].

1.4 Results and Discussion

In this section we present our empirical results. We present the main results, based on (plain and discounted) average market capitalization in the fourth quarter of 2020 in Section 1.4.1. In Section 1.4.2 we then show that we obtain qualitatively similar results when we vary the period over which we measure market capitalization.

1.4.1 Main Results

In Table 1.2 we show the results for the plain market capitalization. Panel A (columns 1-4) show the results of the two-stage regression analysis while Panel B (columns 5-8) shows the LASSO results. We start with the presentation of the two-stage regression results. The four columns of Panel A show the results of the encompassing regression. The corresponding intragroup regression results that determine the set of variables included in the encompassing regression are shown in Appendix 1.6.1. Column 1 (2,3) shows the results that we obtain when all variables with a p-value below 0.1 (0.2, 0.3) in the intragroup regression are included in the encompassing regression. Column 4 shows the results that we obtain when all independent variables are included in the encompassing regression.[33] Note that the F-statistics shown in the last line indicate that the independent variables have significant explanatory power for the market capitalization of the cryptocurrencies only in columns 1, 2 and 3, but not in column 4 where all independent variables are included. This finding supports our choice of the two-stage regression design.

The encompassing regressions of the two-step regression approach yield three main results. First, the age of a cryptocurrency network, as measured by the variable *DaysAge* (the number of days since the launch of the genesis block, rescaled to the range $[0, 1]$), is positively related to market capitalization.[34] Second, spin-offs from other cryptocurrencies (forks) have significantly lower market values (variable *Fork*). This result is in line with our argument that such networks are almost identical copies of already existing cryptocurrencies (the respective parent networks), and that this lack of innovation negatively affects valuation. Third, we find that a configuration of design features similar to that of Bitcoin is associated with higher valuation. Remember that we defined our dummy variables such that they essentially capture departures from the Bitcoin design, and that the continuous variables *HashAge* and *BlockTimeAverage* take on the value zero for the Bitcoin network and positive values for other cryptocurrencies in the sample. The observation that most coefficient signs in Table 1.2 are negative thus implies that deviations from the Bitcoin design result in lower valuation. This result may be due to the first mover advantage of the Bitcoin network, to the higher attention that Bitcoin receives in comparison to other networks, and to the fact that cryptocurrency users and investors are better informed about the details of Bitcoin than about those of its contenders.

We now turn to the discussion of the LASSO results. In Figure 1.1 we graphically illustrate for the first 200 (out of a total of 10,000) simulations the variables which were selected by the procedure and the magnitude of the coefficient estimates. The lines represent the independent variables and the columns the 200 simulation runs. Green (red) color indicates a positive (negative) coefficient estimate, and the intensity of the color represents the magnitude of the estimate. We show numerical results in columns 5 to 8 of Table 1.2. In column 5 we show the frequency with which a variable is selected. In columns 6 and 7 we show, conditional on a variable being selected, the frequency of positive and negative coefficient estimates, respectively. In column 8 we show the mean coefficient estimate.[35]

[33] The number of observations is lower in columns 3 and 4 than in columns 1 and 2 because variables with missing values (such as *BlockTimeAverage*) are included in columns 3 and 4, but not in columns 1 and 2.

[34] The variable *DaysAge* potentially correlates with some other predictors. When we exclude *DaysAge* from the regression model we still observe the same significant effects, and no other variable shows up to be consistently significant.

[35] The means are unconditional, i.e. they are calculated based on all 10,000 simulation runs. Whenever a variable is not selected, the coefficient estimate is set to 0. Conditional means (i.e. means that are calculated conditional on the respective variable being selected by the LASSO procedure) can be obtained by combining

Table 1.2: Market capitalization regression analysis of fourth quarter 2020

	Panel A: Encompassing regression				Panel B: LASSO			
	(1) $p < 0.1$	(2) $p < 0.2$	(3) $p < 0.3$	(4) All	(5) Included	(6) Positive	(7) Negative	(8) ∅ coefficent
Constant	0.133	0.311*	0.367*	0.044	100%	100%	0%	0.167
	(0.112)	(0.159)	(0.216)	(0.486)				
DaysAge	0.583***	0.454**	0.486**	0.724*	80.86%	100%	0%	0.166
	(0.189)	(0.211)	(0.229)	(0.417)				
DeveloperNPO			−0.057	−0.151	0%	-	-	0
			(0.120)	(0.198)				
DeveloperPrivate	−0.066	−0.045	−0.110	−0.241	23.02%	0%	100%	−0.006
	(0.071)	(0.076)	(0.111)	(0.177)				
MajorityChanges				0.021	0%	-	-	0
				(0.128)				
CodeNonC				0.235	2.75%	100%	0%	0.000
				(0.142)				
CodePublic				−0.002	2.75%	100%	0%	0.000
				(0.291)				
Fork	−0.132*	−0.164**	−0.193**	−0.226*	28.19%	0%	100%	−0.025
	(0.068)	(0.072)	(0.077)	(0.124)				
ConsensusPoSDPoS	−0.023	−0.138	−0.079	−0.105	2.03%	0%	100%	−0.000
	(0.083)	(0.123)	(0.139)	(0.200)				
ConsensusOther		−0.150	−0.109	0.051	0%	-	-	0
		(0.126)	(0.147)	(0.218)				
HashAge	−0.153	−0.232	−0.104	0.005	17.38%	0%	100%	−0.003
	(0.121)	(0.140)	(0.168)	(0.271)				
CurveNonECDSA				−0.013	0%	-	-	0
				(0.143)				
MaxSupply				−0.079	0%	-	-	0
				(0.199)				
SupplyCirculation				0.105	0%	-	-	0
				(0.242)				
Deflationary				0.064	0%	-	-	0
				(0.182)				
FixedSupply		−0.051	−0.013	0.001	0%	-	-	0
		(0.086)	(0.095)	(0.170)				
RewardCoinbase			−0.019	0.085	0%	-	-	0
			(0.099)	(0.166)				
RewardAlternative		−0.023	0.040	0.261	0%			0
		(0.084)	(0.116)	(0.213)				
BlockTimeAverage			−0.125	0.013	0.16%	0%	100%	−0.000
			(0.149)	(0.228)				
TransactionFeeObligation				−0.064	0%	-	-	0
				(0.139)				
TipSpecialTreatment				0.040	0%	-	-	0
				(0.116)				
NoFeeTipForMinerForger	0.107	0.126	0.085	0.117	26.19%	100%	0%	0.012
	(0.080)	(0.090)	(0.113)	(0.171)				
IntentionNonPayment				0.284	0%	-	-	0
				(0.245)				
SmartContractSupport				−0.386**	24.24%	0%	100%	−0.014
				(0.187)				
UsageBeyondPayment				−0.061	0%	-	-	0
				(0.139)				
LedgerStyleOther			0.381	0.444	39.35%	100%	0%	0.049
			(0.270)	(0.439)				
AccountingBalance				0.026	0%	-	-	0
				(0.183)				
Anonymous		−0.031	−0.023	−0.048	24.24%	0%	100%	−0.005
		(0.086)	(0.094)	(0.119)				
NonAnonymous			0.373	0.561	25.50%	100%	0%	0.034
			(0.228)	(0.346)				
Observations	68	68	65	59	∅ Observations:			
R^2	0.256	0.294	0.386	0.525	59			
Adjusted R^2	0.182	0.170	0.198	0.082				
F Statistic	3.491***	2.375**	2.055**	1.185	∅ R^2:			
	(df=6;61)	(df=10;57)	(df=15;49)	(df=28;30)	0.108			

This table reports results of the cross-sectional regression of the average market capitalization in the fourth quarter of 2020 on the design feature variables and provides statistics for the variable selection process when applying LASSO with cross-validation. The encompassing models (1), (2), and (3) include the design feature variables with p-values below 0.1, 0.2 and 0.3 in the intra-group regressions, respectively. We control for multicollinearity and find that all variance inflation factors (VIF) in (1) - (3) are below 4.32. Column (4) shows the results for the case that all design feature variable are included (max. VIF of 8.65). Standard errors are given in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Column (5) reports the percentage of cases in which a variable is selected by LASSO while (6) and (7) indicate the related sign of the coefficient. Column (8) reports the average of the parameter estimates indicating the economic significance.

The variable that is most frequently selected (80.0% of the simulations) is the age of a cryptocurrency. Whenever selected the coefficient estimates are, consistent with the results of the two-stage regression approach, positive. All other variables are much less frequently selected. The variable *Fork* is selected in 28.2% of the simulations and the coefficient estimates are, again in line with the regression results presented above, always negative.

The LASSO procedure further selects the variables *DeveloperPrivate* (effect sign: −), *HashAge* (−), *NoFeeTipForMinerForger* (+), *SmartContractSupport* (−), *LedgerStyleOther* (+), *Anonymous* (−), and *NonAnonymous* (+) with reasonable frequency. In all cases the estimated direction of the effect is consistent with the sign of the coefficient estimates in the two-stage regressions.[36] The negative sign of *DeveloperPrivate* indicates that cryptocurrencies which were developed by for-profit entities have lower valuation. The negative impact of *HashAge* on valuation implies that, contrary to our prediction, younger hash functions, which arguably offer higher levels of security, do not increase market capitalization, ceteris paribus.[37] The positive coefficient sign of the variable *NoFeeTipForMinerForger* indicates that networks that do not pass on any transaction fees and/or tips to agents who maintain the integrity of the network have a higher market capitalization. In networks that directly reward contributions to transaction processing with fees and/or tips, transaction fees obviously play an important role. One drawback is that such transaction fees can lead to user non-participation: The fees directly cause some users to drop out, while longer waiting times cause other users who pay fees to drop out as well (Easley et al., 2019; Huberman et al., 2021; Basu et al., 2019). In addition, this can lead to adverse effects related to network security (Pagnotta, 2022). Overall, these fees can increase the vulnerability of the system, which may serve to explain the positive influence of the variable *NoFeeTipForMinerForger*.

The positive coefficient signs for the variable *LedgerStyleOther* indicate that non-blockchain-based cryptocurrencies have higher market valuation. This finding should be interpreted with care, though, because our sample only contains five cryptocurrencies with that feature. The effect signs of the variables *Anonymous* and *NonAnonymous* imply that cryptocurrencies that allow completely anonymous transactions have lower market values while those that require disclosure of real-world identities have higher market values. The former result may be due to concerns that fully anonymous networks might be misused for illegal transactions. The latter result may reflect the expectation of regulatory acceptance of non-anonymous cryptocurrencies.

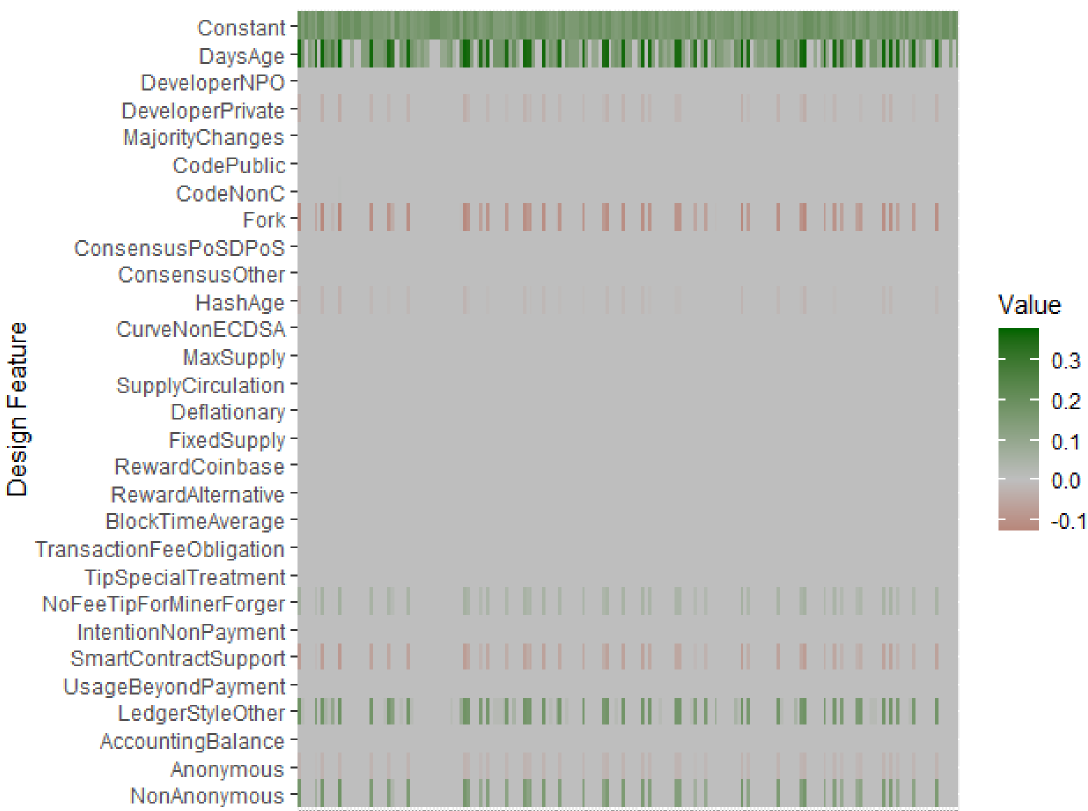
The negative effect on market valuation ascribed to the variable *SmartContractSupport* runs counter to the intuition that a network that allows for smart contracts allows alternative uses beyond making payments and should thus be more valuable. However, smart contracts may also be gateways for fraudulent behavior and/or may be subject to coding errors which might result in security breaches.

the unconditional means with the data on the selection frequency provided in column 5 of the table.

[36] Note, though, that the coefficient estimates were insignificant in the encompassing regression. However, the coefficients of the variables *DeveloperPrivate*, *HashAge*, and *NoFeeTipForMinerForger* were significant at the 10% level or better in the intra-group regressions.

[37] Remember that we defined the variable *HashAge* such that larger values mean younger hash functions.

Figure 1.1: LASSO variable selection and economic magnitudes (Marketcap Q4 2020)



This figure shows the economic magnitude of the estimated coefficients for each design feature covering 200 randomly selected training and validation data subset compositions from our LASSO approach. Red (green) bars refer to a negative (positive) coefficient estimate, while grey bars refer to coefficient estimates equal to zero, i.e. to non-selected design features. More intense colors refer to stronger economic magnitude.

Table 1.3: Discounted market capitalization regression analysis of fourth quarter 2020

	Panel A: Encompassing regression				Panel B: LASSO			
	(1) $p < 0.1$	(2) $p < 0.2$	(3) $p < 0.3$	(4) All	(5) Included	(6) Positive	(7) Negative	(8) ∅ coefficent
Constant	0.168**	0.186**	0.436***	0.000	100%	100%	0%	0.189
	(0.064)	(0.068)	(0.144)	(0.469)				
DaysAge	0.101	0.100	0.016	0.175	0%	-	-	0
	(0.158)	(0.158)	(0.169)	(0.402)				
DeveloperNPO		−0.059	−0.117	−0.232	0%	-	-	0
		(0.070)	(0.103)	(0.191)				
DeveloperPrivate			−0.150	−0.303*	0%	-	-	0
			(0.093)	(0.170)				
MajorityChanges				−0.044	0%	-	-	0
				(0.123)				
CodeNonC				0.179	0%	-	-	0
				(0.137)				
CodePublic				0.094	0%	-	-	0
				(0.280)				
Fork	−0.170***	−0.175***	−0.221***	−0.315**	80.52%	0%	100%	−0.066
	(0.062)	(0.063)	(0.067)	(0.120)				
ConsensusPoSDPoS				−0.002	0%	-	-	0
				(0.193)				
ConsensusOther				0.116	0%	-	-	0
				(0.210)				
HashAge				0.043	0%	-	-	0
				(0.261)				
CurveNonECDSA				0.056	0%	-	-	0
				(0.138)				
MaxSupply				−0.091	0%	-	-	0
				(0.192)				
SupplyCirculation				0.189	0%	-	-	0
				(0.234)				
Deflationary				0.032	0%	-	-	0
				(0.176)				
FixedSupply				−0.008	0%	-	-	0
				(0.164)				
RewardCoinbase				0.225	0%	-	-	0
				(0.160)				
RewardAlternative				0.329	0%	-	-	0
				(0.205)				
BlockTimeAverage			−0.177	−0.010	0%	-	-	0
			(0.123)	(0.220)				
TransactionFeeObligation				−0.069	0%	-	-	0
				(0.134)				
TipSpecialTreatment				0.091	0%	-	-	0
				(0.112)				
NoFeeTipForMinerForger	0.077	0.075	0.169**	0.199	73.74%	100%	0%	0.018
	(0.075)	(0.075)	(0.083)	(0.165)				
IntentionNonPayment				0.156	0%	-	-	0
				(0.236)				
SmartContractSupport				−0.261	0%	-	-	0
				(0.180)				
UsageBeyondPayment			0.062	0.002	0%	-	-	0
			(0.072)	(0.134)				
LedgerStyleOther				0.077	0%	-	-	0
				(0.423)				
AccountingBalance				−0.0001	0%	-	-	0
				(0.176)				
Anonymous				−0.041	0%	-	-	0
				(0.115)				
NonAnonymous	0.400**	0.415**	0.368*	0.611*	65.97%	100%	0%	0.036
	(0.184)	(0.186)	(0.186)	(0.333)				
Observations	68	68	65	59	∅ Observations:			
R^2	0.187	0.196	0.291	0.478	59			
Adjusted R^2	0.135	0.131	0.190	−0.009				
F Statistic	3.625**	3.028**	2.878***	0.982	∅ R^2:			
	(df=4;63)	(df=5;62)	(df=8;56)	(df=28;30)	0.087			

This table reports results of the cross-sectional regression of the average discounted market capitalization in the fourth quarter of 2020 on the design feature variables and provides statistics for the variable selection process when applying LASSO with cross-validation. The encompassing models (1), (2), and (3) include the design feature variables with p-values below 0.1, 0.2, and 0.3 in the intra-group regressions, respectively. We control for multicollinearity and find that all variance inflation factors (VIF) in (1) - (3) are below 2.27. Column (4) shows the results for the case that all design feature variable are included (max. VIF of 8.65). Standard errors are given in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. Column (5) reports the percentage of cases in which a variable is selected by LASSO while (6) and (7) indicate the related sign of the coefficient. Column (8) reports the average of the parameter estimate indicating the economic significance.

We next turn to the results for the discounted market capitalization. The dependent variable is the time-series average of the discounted market capitalization (equation (1.1)) during the last quarter of 2020. Otherwise the analysis is identical to the one presented above. We present in Table 1.3 results of the two-stage regressions (columns 1-4) and the LASSO results (columns 5-8). In addition, a graphical representation of the results for the first 200 runs of the LASSO procedure can be found in Figure 1.2.

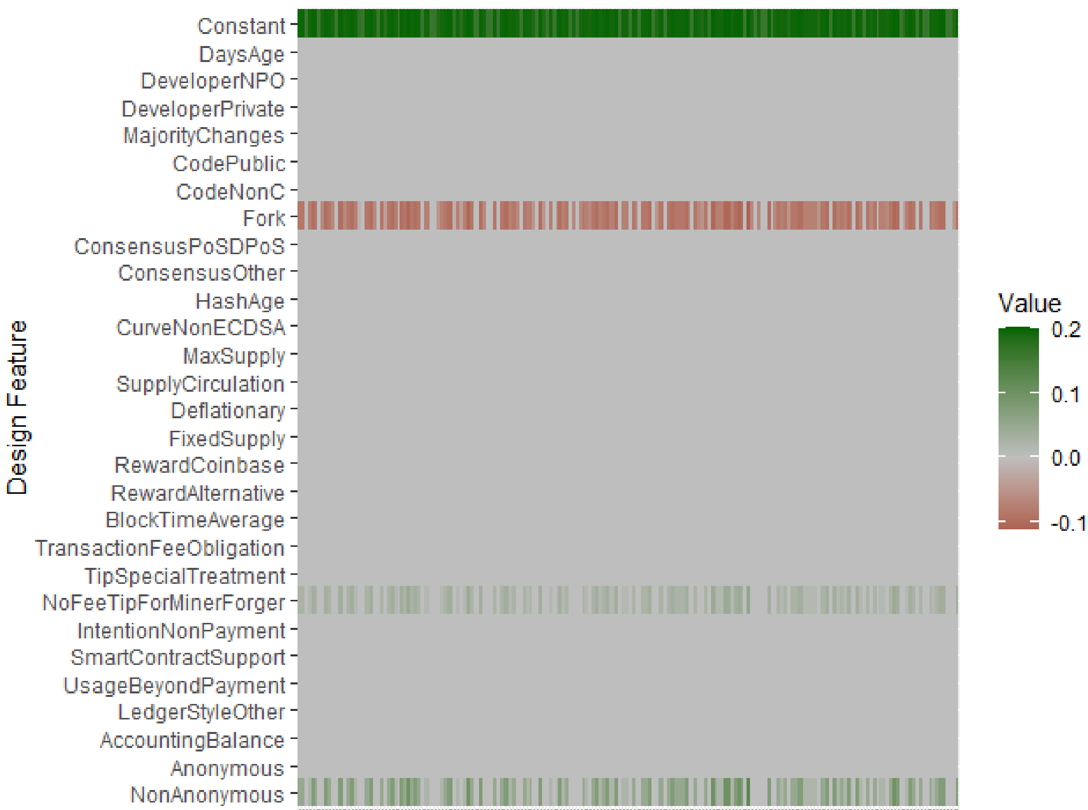
The coefficient estimates for the age of the cryptocurrency (variable *DaysAge*) in the encompassing regressions are much smaller than before and are always insignificant. Furthermore, the variable is never selected by the LASSO approach. These results indicate that the discounting procedure successfully removed the effect of age on market valuation.

Regarding the influence of the design features on the market valuation, the results for discounted market capitalization are similar to those for plain market capitalization. We note, though, that the LASSO procedure selects fewer variables when we use discounted market capitalization as the dependent variable. As before, we find that spin-offs from other cryptocurrencies (variable *Fork*) have lower valuation. The respective coefficient estimate is negative and highly significant in the encompassing regression, and it is very frequently (80.5%) selected by the LASSO procedure, always with a negative coefficient estimate. Finally, and again consistent with our previous results, we obtain a positive coefficient estimate for the variable *NonAnonymous*. It implies that cryptocurrency networks that require disclosure of real world identities tend to have higher valuation. Furthermore, our earlier result that networks in which agents who verify transactions are rewarded by a scheme independent of fees and/or tips have a higher market valuation is also confirmed. The respective coefficient estimate (variable *NoFeeTipForMinerForger*) in the encompassing regressions is always positive (significantly so in one case), and the variable is frequently selected (73.7%) by the LASSO procedure, always with a positive coefficient sign. Finally, there is still (albeit weak and only based on the two-stage regression analysis) evidence that cryptocurrencies developed by private for-profit entities are less valuable. We no longer find evidence that younger hash functions are associated with lower valuation, nor is there evidence that fully anonymous networks are less valuable.

1.4.2 Robustness

So far we have analyzed whether design features can explain the average (plain and discounted) market capitalization in the fourth quarter of 2020. While averaging over values for an entire quarter should make our results insensitive to day-to-day fluctuations in cryptocurrency prices, we still have to establish that our findings are not specific to the single quarter we have considered. To this end we repeat our entire analysis using the average (plain and discounted) market capitalization over (1) the entire year 2020 and (2) the first, second and third quarter of 2020. The results for the full year are shown in Tables 1.6, 1.7, 1.8, and 1.9 in Appendix 1.6.2. The results for quarters 1 to 3 are qualitatively similar to those reported in the paper and are omitted.

Figure 1.2: LASSO variable selection and economic magnitudes (Discounted marketcap Q4 2020)



This figure shows the economic magnitude of the estimated coefficients for each design feature covering 200 randomly selected training and validation data subset compositions from our LASSO approach. Red (green) bars refer to a negative (positive) coefficient estimate, while grey bars refer to coefficient estimates equal to zero, i.e. to non-selected design features. More intense colors refer to stronger economic magnitude.

The two-stage regression approach for the plain market capitalization averaged over the full year (Table 1.6) fully confirms the three main results highlighted previously. Older cryptocurrencies have higher market valuation, forks have lower market capitalization, and deviations from the Bitcoin design are associated with lower market capitalization. The latter conclusion, as before, follows from the fact that the overwhelming majority of the coefficients of the encompassing regression are negative, and that we have defined all independent variables such that their values for the Bitcoin network are zero. The results in Table 1.6 also confirm our previous finding that networks where the reward of agents who verify transactions are independent of fees and/or tips have higher valuation. The LASSO results in Table 1.7 are fully consistent with those just discussed. Furthermore, they are also consistent with the LASSO results in Table 1.2 in that the variables *DeveloperPrivate* (effect sign: −), *HashAge* (−), *SmartContractSupport* (−), *LedgerStyleOther* (+), *Anonymous* (−), and *NonAnonymous* (+) are again selected with reasonable frequencies, and have the same coefficient signs as in Table 1.2.

The results for the discounted market valuation, averaged over the entire year 2020 (Tables 1.8 and 1.9) again support all previous conclusions. The age of a cryptocurrency does not significantly affect its discounted market capitalization, forks have lower valuation, and most coefficient estimates in the encompassing regressions are negative, implying that deviations from the Bitcoin design are associated with lower valuation. Furthermore, the result that non-anonymous networks have higher value is confirmed, as is the earlier result that networks in which transactions fees and/or tips are passed to agents maintaining the network's integrity at all have higher market capitalization. Even the (weak) evidence that cryptocurrencies developed by private for-profit entities are less valuable is confirmed.

1.5 Conclusion

In this paper we analyze whether the value of cryptocurrencies as measured by their market capitalization can be related to specific cryptocurrency design features. To this end we first propose a taxonomy of cryptocurrency design features and hand-collect a dataset that contains these features for a set of 79 cryptocurrencies. We then use two different methodological approaches, a two-stage regression analysis in the tradition of Karnaukh et al. (2015) and LASSO regressions, to analyze whether any of these design features are cross-sectionally related to cryptocurrency valuation. To control for the potential effect of the age of a cryptocurrency on its value we repeat the analysis using discounted instead of plain market capitalization as our dependent variable.

We find that cryptocurrencies that were spun off other cryptocurrencies (forks) are less valuable. On the other hand, cryptocurrencies where agents who verify transactions are rewarded by a scheme independent of fees and/or tips tend to be more valuable. Interestingly, cryptocurrencies that require the disclosure of the real-world identities of its users have higher values, possibly in expectation of regulatory approval of these networks. Apart from that we find that deviations from the design of Bitcoin tend to be associated with lower valuation.

Thus, even though Bitcoin may not be the most technologically advanced cryptocurrency, users and investors apparently value its design.

Overall, we provide evidence that design features partly affect the market valuation of cryptocurrencies. Due to the relatively new underlying technology of cryptocurrencies and its complexity, investors might not be aware of crucial design feature differences between the different cryptocurrency networks. Thus, they might not value the technology per se, but rather the hope to invest in the "next Bitcoin".

While we consider the impact of a large number of design features on cryptocurrency valuation, we do not take into account interactions between different design features. Such interactions may be relevant, though. For instance, the influence of shorter blocktime in a PoS network is expected to be positive due to higher throughput enabled by shorter blocktimes. In contrast, if the blocktime is too small in a PoW network, attacks on the network by fraudulent agents may become more likely which, in turn, may result in more reluctant network adoption and eventually in reduced market capitalization. Extending our research approach to incorporate such interaction effects is a promising avenue for future research.

1.6 Appendix to Chapter 1

1.6.1 Results of the Intra-group Regressions in the Main Analysis

Table 1.4: Intra-group market capitalization regressions of fourth quarter 2020

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.246	0.397***	0.311**	0.355**	0.270***	0.204***
	(0.312)	(0.000)	(0.013)	(0.029)	(0.000)	(0.001)
DeveloperNPO	−0.145					
	(0.224)					
DeveloperPrivate	−0.190*					
	(0.074)					
MajorityChanges	−0.024					
	(0.772)					
CodeNonC	−0.025					
	(0.769)					
CodePublic	0.185					
	(0.397)					
Fork	−0.183**					
	(0.024)					
ConsensusPoSDPoS		−0.208**				
		(0.039)				
ConsensusOther		−0.150				
		(0.179)				
HashAge		−0.280*				
		(0.056)				
CurveNonECDSA		−0.022				
		(0.766)				
MaxSupply			0.011			
			(0.899)			
SupplyCirculation			−0.0003			
			(0.917)			
Deflationary			0.058			
			(0.634)			
FixedSupply			−0.136			
			(0.166)			
RewardCoinbase			−0.118			
			(0.231)			
RewardAlternative			−0.126			
			(0.184)			
BlockTimeAverage				−0.199		
				(0.206)		
TransactionFeeObligation				−0.015		
				(0.873)		
TipSpecialTreatment				−0.030		
				(0.745)		
NoFeeTipForMinerForger				0.191*		
				(0.089)		
IntentionNonPayment					−0.032	
					(0.801)	
SmartContractSupport					−0.113	
					(0.341)	
UsageBeyondPayment					0.007	
					(0.931)	
LedgerStyleOther						0.161
						(0.248)
AccountingBalance						−0.037
						(0.622)
Anonymous						−0.111
						(0.179)
NonAnonymous						0.270
						(0.214)
Observations	68	68	68	59	68	68
R^2	0.109	0.089	0.077	0.078	0.047	0.073
Adjusted R^2	0.022	0.031	−0.014	0.009	0.003	0.014
F Statistic	1.248	1.533	0.845	1.138	1.060	1.245
	(df=6;61)	(df=4;63)	(df=6;61)	(df=4;54)	(df=3;64)	(df=4;63)

This table reports results of the cross-sectional intra-group regression of the average market capitalization in the fourth quarter of the year 2020 on the design feature variables.We control for multicollinearity and find that all variance inflation factors (VIF) are below 2.6. p-Values are given in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 1.5: Intra-group discounted market capitalization regressions of fourth quarter 2020

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.196	0.202**	0.158	0.269*	0.116*	0.125**
	(0.368)	(0.033)	(0.170)	(0.068)	(0.066)	(0.026)
DeveloperNPO	−0.149					
	(0.163)					
DeveloperPrivate	−0.119					
	(0.209)					
MajorityChanges	−0.015					
	(0.837)					
CodeNonC	0.037					
	(0.621)					
CodePublic	0.139					
	(0.476)					
Fork	−0.210***					
	(0.004)					
ConsensusPoSDPoS		0.018				
		(0.846)				
ConsensusOther		−0.066				
		(0.525)				
HashAge		−0.111				
		(0.415)				
CurveNonECDSA		−0.046				
		(0.510)				
MaxSupply			0.001			
			(0.993)			
SupplyCirculation			0.0002			
			(0.927)			
Deflationary			0.007			
			(0.950)			
FixedSupply			−0.094			
			(0.306)			
RewardCoinbase			−0.014			
			(0.882)			
RewardAlternative			0.054			
			(0.543)			
BlockTimeAverage				−0.150		
				(0.297)		
TransactionFeeObligation				−0.006		
				(0.948)		
TipSpecialTreatment				−0.028		
				(0.740)		
NoFeeTipForMinerForger				0.210**		
				(0.042)		
IntentionNonPayment					0.089	
					(0.442)	
SmartContractSupport					−0.102	
					(0.349)	
UsageBeyondPayment					0.085	
					(0.245)	
LedgerStyleOther						−0.096
						(0.446)
AccountingBalance						0.057
						(0.396)
Anonymous						−0.056
						(0.456)
NonAnonymous						0.349*
						(0.078)
Observations	68	68	68	59	68	68
R^2	0.144	0.041	0.032	0.091	0.041	0.089
Adjusted R^2	0.060	−0.020	−0.064	0.024	−0.004	0.031
F Statistic	1.710	0.669	0.333	1.349	0.905	1.540
	(df=6;61)	(df=4;63)	(df=6;61)	(df=4;54)	(df=3;64)	(df=4;63)

This table reports results of the cross-sectional intra-group regression of the average discounted market capitalization in the fourth quarter of the year 2020 on the design feature variables.We control for multicollinearity and find that all variance inflation factors (VIF) are below 2.6. p-Values are given in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

1.6.2 Results of the Robustness Analysis

Table 1.6: Market capitalization regression analysis of year 2020

	(1)	(2)	(3)	(4)	(5)	(6)	(7) $p < 0.1$	(8) $p < 0.2$	(9) $p < 0.3$	(10)
Constant	0.216	0.373***	0.273**	0.328**	0.234***	0.172***	0.119	0.382**	0.497***	0.114
	(0.219)	(0.089)	(0.109)	(0.143)	(0.060)	(0.054)	(0.101)	(0.168)	(0.180)	(0.425)
DaysAge							0.516***	0.419**	0.282	0.655*
							(0.171)	(0.183)	(0.201)	(0.355)
DeveloperNPO	−0.150							−0.022	−0.029	−0.159
	(0.106)							(0.102)	(0.104)	(0.170)
DeveloperPrivate	−0.172*						−0.043	−0.080	−0.089	−0.235
	(0.094)						(0.064)	(0.097)	(0.099)	(0.152)
MajorityChanges	−0.008									0.037
	(0.074)									(0.108)
CodeNonC	−0.014									0.249*
	(0.076)									(0.123)
CodePublic	0.158									−0.072
	(0.197)									(0.253)
Fork	−0.170**						−0.115*	−0.175**	−0.201***	−0.191*
	(0.071)						(0.061)	(0.067)	(0.068)	(0.101)
ConsensusPoSDPoS		−0.198**					−0.030	−0.103	−0.081	−0.133
		(0.088)					(0.076)	(0.125)	(0.126)	(0.173)
ConsensusOther		−0.142						−0.109	−0.134	0.009
		(0.098)						(0.128)	(0.132)	(0.187)
HashAge		−0.280**					−0.155	−0.205	−0.215	−0.041
		(0.130)					(0.110)	(0.138)	(0.147)	(0.233)
CurveNonECDSA		−0.031								−0.046
		(0.066)								(0.120)
MaxSupply			−0.002							−0.026
			(0.079)							(0.169)
SupplyCirculation			−0.0002							0.043
			(0.003)							(0.207)
Deflationary			0.047							0.034
			(0.111)							(0.158)
FixedSupply			−0.129					−0.021	−0.008	−0.033
			(0.088)					(0.083)	(0.087)	(0.148)
RewardCoinbase			−0.090							0.082
			(0.086)							(0.131)
RewardAlternative			−0.103						0.008	0.233
			(0.084)						(0.081)	(0.171)
BlockTimeAverage				−0.202				−0.130	−0.113	−0.024
				(0.141)				(0.128)	(0.133)	(0.192)
TransactionFeeObligation				−0.029						−0.084
				(0.087)						(0.120)
TipSpecialTreatment				−0.007						0.056
				(0.081)						(0.101)
NoFeeTipForMinerForger				0.199*			0.090	0.177**	0.119	0.125
				(0.100)			(0.072)	(0.083)	(0.097)	(0.143)
IntentionNonPayment					−0.011					0.355
					(0.107)					(0.210)
SmartContractSupport					−0.121				−0.115	−0.433**
					(0.101)				(0.085)	(0.162)
UsageBeyondPayment					0.011					−0.039
					(0.070)					(0.117)
LedgerStyleOther						0.124			0.319	0.496
						(0.118)			(0.226)	(0.375)
AccountingBalance						−0.021				0.070
						(0.067)				(0.157)
Anonymous						−0.094			−0.003	−0.026
						(0.075)			(0.082)	(0.098)
NonAnonymous						0.187				0.405
						(0.198)				(0.285)
Observations	71	71	71	61	71	71	71	67	67	61
R^2	0.107	0.103	0.067	0.099	0.049	0.054	0.245	0.331	0.383	0.564
Adjusted R^2	0.023	0.048	−0.021	0.035	0.006	−0.003	0.174	0.212	0.218	0.183
F Statistic	1.279	1.887	0.763	1.542	1.146	0.948	3.452***	2.776***	2.310**	1.479
	(df=6;64)	(df=4;66)	(df=6;64)	(df=4;56)	(df=3;67)	(df=4;66)	(df=6;64)	(df=10;56)	(df=14;52)	(df=28;32)

This table reports results of the cross-sectional regression of the average market capitalization in the whole year 2020 on the design feature variables. Columns (1) - (6) shows the coefficients for the intra-group regressions. Models (7), (8), and (9) include the design feature variables with intra-group regression p-values below 0.1, 0.2, and 0.3, respectively. We control for multicollinearity and find that all variance inflation factors (VIF) in (1) - (7) are below 2.4 and below 4.43 in (8) and (9). Column (10) shows the results for the case that all design feature variable are included (max. VIF of 8.78). Standard errors are given in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 1.7: LASSO variable selection for market capitalization regression of year 2020

	(1) Included	(2) Positive	(3) Negative	(4) ∅ coefficent
Constant	100%	100%	0%	0.162
DaysAge	51.16%	100%	0%	0.082
DeveloperNPO	0%	-	-	0
DeveloperPrivate	13.27%	0%	100%	-0.002
MajorityChanges	0%	-	-	0
CodeNonC	1.78%	100%	0%	0.000
CodePublic	0%	-	-	0
Fork	16.93%	0%	100%	-0.015
ConsensusPoSDPoS	0.27%	0%	100%	-0.000
ConsensusOther	0%	-	-	0
HashAge	14.95%	0%	100%	-0.007
CurveNonECDSA	0%	-	-	0
MaxSupply	0%	-	-	0
SupplyCirculation	0%	-	-	0
Deflationary	0%	-	-	0
FixedSupply	1.78%	0%	100%	-0.000
RewardCoinbase	0%	-	-	0
RewardAlternative	0%			0
BlockTimeAverage	10.82%	0%	100%	-0.002
TransactionFeeObligation	0.01%	0%	100%	-0.000
TipSpecialTreatment	0%	-	-	0
NoFeeTipForMinerForger	16.93%	100%	0%	0.009
IntentionNonPayment	0%	-	-	0
SmartContractSupport	16.03%	0%	100%	-0.010
UsageBeyondPayment	0%	-	-	0
LedgerStyleOther	19.80%	100%	0%	0.024
AccountingBalance	0.015%	100%	0%	0.000
Anonymous	14.95%	0%	100%	-0.002
NonAnonymous	14.31%	100%	0%	0.009
∅ Observations			61	
∅ Fraction of (null) deviance explained			0.062	

This table provides statistics for the variable selection process when applying LASSO with cross-validation using the average market capitalization in the whole year 2020 as the dependent variable. Column (1) reports the percentage of cases in which a variable is selected by LASSO while (2) and (3) indicate the related sign of the coefficient. Column (4) reports the average of the parameter estimate indicating the economic significance. Deviance is defined as $2\,(loglike_{sat} - loglike)$, where $loglike_{sat}$ is the log-likelihood for the saturated model. Null deviance is defined to be $2\,(loglike_{sat} - NULL)$ with $NULL$ referring to the intercept model.

Table 1.8: Discounted market capitalization regression analysis of year 2020

	(1)	(2)	(3)	(4)	(5)	(6)	(7) $p < 0.1$	(8) $p < 0.2$	(9) $p < 0.3$	(10)
Constant	0.225	0.218**	0.151	0.298*	0.127**	0.129**	0.203***	0.361***	0.527***	0.136
	(0.228)	(0.096)	(0.117)	(0.152)	(0.063)	(0.056)	(0.067)	(0.110)	(0.153)	(0.501)
DaysAge							0.028	−0.053	−0.078	0.050
							(0.167)	(0.173)	(0.179)	(0.419)
DeveloperNPO	−0.164							−0.170*	−0.116	−0.214
	(0.111)							(0.098)	(0.109)	(0.200)
DeveloperPrivate	−0.133							−0.149	−0.165*	−0.329*
	(0.098)							(0.091)	(0.097)	(0.180)
MajorityChanges	−0.010									−0.039
	(0.077)									(0.128)
CodeNonC	0.058									0.194
	(0.079)									(0.144)
CodePublic	0.109									0.053
	(0.205)									(0.298)
Fork	−0.198***						−0.159**	−0.206***	−0.225***	−0.296**
	(0.074)						(0.066)	(0.070)	(0.072)	(0.119)
ConsensusPoSDPoS		0.018								−0.012
		(0.096)								(0.204)
ConsensusOther		−0.063								0.108
		(0.106)								(0.221)
HashAge		−0.111								−0.010
		(0.141)								(0.275)
CurveNonECDSA		−0.052								0.039
		(0.072)								(0.141)
MaxSupply			0.004							−0.061
			(0.085)							(0.199)
SupplyCirculation			0.0001							0.178
			(0.003)							(0.244)
Deflationary			−0.0001							0.004
			(0.119)							(0.187)
FixedSupply			−0.089							−0.021
			(0.094)							(0.175)
RewardCoinbase			0.010							0.209
			(0.093)							(0.154)
RewardAlternative			0.060							0.276
			(0.089)							(0.201)
BlockTimeAverage				−0.160					−0.219*	−0.069
				(0.150)					(0.131)	(0.226)
TransactionFeeObligation				−0.019						−0.083
				(0.092)						(0.142)
TipSpecialTreatment				−0.014						0.115
				(0.086)						(0.118)
NoFeeTipForMinerForger				0.201*			0.045	0.068	0.160*	0.240
				(0.106)			(0.078)	(0.079)	(0.088)	(0.168)
IntentionNonPayment					0.067					0.137
					(0.113)					(0.247)
SmartContractSupport					−0.078					−0.269
					(0.107)					(0.191)
UsageBeyondPayment					0.090				0.060	0.031
					(0.074)				(0.076)	(0.138)
LedgerStyleOther						−0.125				0.036
						(0.122)				(0.442)
AccountingBalance						0.067				0.046
						(0.069)				(0.185)
Anonymous						−0.030				0.012
						(0.077)				(0.116)
NonAnonymous						0.355*	0.402**	0.423**	0.363*	0.521
						(0.204)	(0.198)	(0.196)	(0.199)	(0.336)
Observations	71	71	71	61	71	71	71	71	67	61
R^2	0.123	0.038	0.028	0.078	0.033	0.083	0.146	0.189	0.263	0.452
Adjusted R^2	0.041	−0.020	−0.063	0.012	−0.010	0.027	0.094	0.113	0.161	−0.027
F Statistic	1.493	0.651	0.306	1.188	0.764	1.485	2.810**	2.482**	2.587**	0.943
	(df=6;64)	(df=4;66)	(df=6;64)	(d =4;56)	(df=3;67)	(df=4;66)	(df=4;66)	(df=6;64)	(df=8;58)	(df=28;32)

This table reports results of the cross-sectional regression of the average market capitalization in the whole year 2020 on the design feature variables. Columns (1) - (6) shows the coefficients for the intra-group regressions. Models (7), (8), and (9) include the design feature variables with intra-group regression p-values below 0.1, 0.2, and 0.3, respectively. We control for multicollinearity and find that all variance inflation factors (VIF) in (1) - (9) are below 2.4. Column (10) shows the results for the case that all design feature variable are included (max. VIF of 8.78). Standard errors are given in parentheses. *, **, and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 1.9: LASSO variable selection for discounted market capitalization regression of year 2020

	(1) Included	(2) Positive	(3) Negative	(4) ∅ coefficent
Constant	100%	100%	0%	0.194
DaysAge	0%	-	-	0
DeveloperNPO	0%	-	-	0
DeveloperPrivate	0%	-	-	0
MajorityChanges	0%	-	-	0
CodeNonC	0%	-	-	0
CodePublic	0%	-	-	0
Fork	65.01%	0%	100%	-0.035
ConsensusPoSDPoS	0%	-	-	0
ConsensusOther	0%	-	-	0
HashAge	0%	-	-	0
CurveNonECDSA	0%	-	-	0
MaxSupply	0%	-	-	0
SupplyCirculation	0%	-	-	0
Deflationary	0%	-	-	0
FixedSupply	0%	-	-	0
RewardCoinbase	0%	-	-	0
RewardAlternative	0%			0
BlockTimeAverage	0%	-	-	0
TransactionFeeObligation	0%	-	-	0
TipSpecialTreatment	0%	-	-	0
NoFeeTipForMinerForger	14.55%	100%	0%	0.001
IntentionNonPayment	0%	-	-	0
SmartContractSupport	0%	-	-	0
UsageBeyondPayment	0%	-	-	0
LedgerStyleOther	0%	-	-	0
AccountingBalance	0%	-	-	0
Anonymous	0%	-	-	0
NonAnonymous	30.03%	100%	0%	0.007
∅ Observations			61	
∅ Fraction of (null) deviance explained			0.035	

This table provides statistics for the variable selection process when applying LASSO with cross-validation using the average discounted market capitalization in the whole year 2020 as the dependent variable. Column (1) reports the percentage of cases in which a variable is selected by LASSO while (2) and (3) indicate the related sign of the coefficient. Column (4) reports the average of the parameter estimate indicating the economic significance. Deviance is defined as $2\,(loglike_{sat} - loglike)$, where $loglike_{sat}$ is the log-likelihood for the saturated model. Null deviance is defined to be $2\,(loglike_{sat} - NULL)$ with $NULL$ referring to the intercept model.

Chapter 2

Design and Volatility of Cryptocurrencies[1]

2.1 Introduction

High volatility appears to be a general characteristic of cryptocurrencies. However, not all cryptocurrencies are equally volatile. Rather, as will be documented below, there are large cross-sectional differences. Understanding the determinants of these volatility differences is important for cryptocurrency investors, regulators, and cryptocurrency developers. In this paper we analyze whether differences in volatility can be traced back to differences in cryptocurrency design. If that were the case, investors could predict the volatility of a cryptocurrency based on its constellation of design features, and developers could deliberately design cryptocurrencies that can be expected to have low volatility. We adopt the taxonomy proposed in Chapter 1, which identifies a wide variety of cryptocurrency design features and sort them into six categories, namely "development", "technical", "supply", "transactions", "usability", and "general" (see Table 1.1 for details). We collect a complete record of these design features for a broad sample of cryptocurrencies and then use LASSO regressions to identify those design features that affect volatility.

Our paper contributes to the literature on cryptocurrency volatility. While many papers investigate the volatility of Bitcoin (e.g. Urquhart (2017), Bystroem and Krygier (2018), Conrad et al. (2018), Baur and Dimpfl (2021)), others examine volatility dynamics across cryptocurrencies and spillover effects (e.g. Yi et al. (2018), Katsiampa et al. (2019)). Some papers further analyze factors and/or statistical models that can explain and predict cryptocurrency volatility (e.g. Baur and Dimpfl (2018), Bouri et al. (2019), Katsiampa (2019), Yen and Cheng (2021), Catania and Grassi (2022), D'Amato et al. (2022)). Our paper

[1]This work is the result of collaboration with Prof. Dr. Erik Theissen, Fabian Eska, and his supervisor Prof. Dr. Marliese Uhrig-Homburg. As a group of co-authors, we thank the participants of the Ghent Workshop on Fintech 2023 in Ghent, Belgium, for valuable comments and discussions. We thank Deutsche Forschungsgemeinschaft (DFG) for financial support under grants TH 724/7-1 and UH 107/5-1.

is, to the best of our knowledge, the first paper that relates cryptocurrency volatility to cryptocurrency design features. The existence of such a relation has been established for cryptocurrency valuation by Hayes (2017) and Chapter 1. Both papers provide evidence that design features affect the market valuation of cryptocurrencies.

Our results are somewhat ambiguous because they partly depend on the measure of volatility (return standard deviation or interquartile range) and the sample period (2020, 2021) that we use. However, some consistent findings emerge. In line with the theoretical model of Bolt and van Oordt (2020), older cryptocurrencies tend to be less volatile. Cryptocurrencies created for purposes other than being a means of payment are more volatile. Further, cryptocurrency volatility tends to be higher for those currencies that were developed by private for-profit entities. We do not find convincing support for the prediction, made by Saleh (2018), that proof-of-work cryptocurrencies are more volatile than those using other consensus mechanisms.

The remainder of the paper is organized as follows. In Section 2.2 we describe our data and methodology, in Section 2.3 we present the results and Section 2.4 concludes.

2.2 Data and Methodology

Cryptocurrencies are characterized by a set of design features. Most of these features are the result of choices that the developers of the cryptocurrencies make. The choice of a consensus mechanism (proof-of-work, proof-of-stake, or others) is a case in point. Other features relate to the development process itself (e.g. is the developer a for-profit organization?). There are various attempts at categorizing design features (e.g. Garriga et al. (2020), Cousins et al. (2019), and Chapter 1). We adopt the taxonomy proposed in Chapter 1, which was designed for an analysis of the relation between cryptocurrency design features and market valuation, a research question that is related to this one. Table 1.1 shows the six categories of design features, lists the variables within each category, and shows their respective definitions. Our dataset contains design feature information, sourced from official network websites, white papers and other reliable sources,[2] for a total of 58 cryptocurrencies.[3]

Besides data on design features we obtained daily data on the returns of all cryptocurrencies in our sample from the exchange APIs and from Kaiko. Our primary data source is the respective exchange API. We use the data obtained from Kaiko to fill in missing values. The data covers ten different cryptocurrency trading venues.[4] The data refers to the respective cryptocurrency traded either against Bitcoin (BTC) or against the US dollar. One

[2] See Section 1.2.2 for details on the data collection procedure.

[3] We initially collected data on design features for 79 cryptocurrencies. However, not all design features are available for all cryptocurrencies. The final sample size is further reduced (to 58 cryptocurrencies) because we require a sufficiently long time series of daily closing prices to be available. Details are provided below.

[4] The venues are Binance, Bitfinex, Kraken, Bitstamp, Coinbase, bitFlyer, Gemini, itBit, Bittrex, and Poloniex. Note that all these exchanges are reliable according to Härdle et al. (2020). A trading venue is considered as reliable if it does not report inflated trading volume.

exchange (Binance) trades cryptocurrencies only against the Euro, not against the US dollar. Therefore, we use Euro prices instead of USD prices for Binance and convert them into USD prices using the daily USD-EUR exchange rate.[5] This procedure leaves us with two sets of prices, cryptocurrencies against USD (referred to as *direct prices* in the sequel) and against BTC. We have BTC prices for all cryptocurrencies in our sample (except BTC itself, of course). We use daily closing prices (defined as the last transaction prices before 00:00 UTC) to calculate daily returns from BTC-denominated prices. These returns comprise our first sample, denoted as the BTC sample. Unfortunately, we do not have USD prices for all cryptocurrencies because some of the currencies are simply not traded against USD on the trading venues from which we source our data.[6] We therefore convert BTC prices into USD prices using the USD-BTC exchange rate of the trading venue under consideration. We refer to these converted USD prices as *indirect prices*. We then compile our second sample, denoted as the USD sample, as follows.

- For those cryptocurrencies for which we only have BTC prices from a trading venue we use the indirect prices to calculate daily returns.
- For those venues where we have BTC and USD prices we calculate daily returns from both direct (USD) and indirect prices and then calculate the volume-weighted average.

Both the BTC sample and the USD sample have distinct advantages and disadvantages. The BTC sample is consistent in that it does not require any currency conversion. However, it is based on transactions of one cryptocurrency against another (BTC) rather than on transactions of a cryptocurrency against a fiat currency. The USD sample has the advantage of measuring prices against a fiat currency but has the disadvantage of being partially based on indirect prices.[7] To assure the robustness of the results we perform our entire analysis for both samples. A notable difference between the two samples is that the USD sample contains Bitcoin while the BTC sample (where Bitcoin is the numeraire) does not. We re-estimated all models for the USD sample after excluding Bitcoin and obtained results similar to those presented in the paper.

We use two measures of volatility, the standard deviation and the interquartile range of the daily returns. To aggregate the data from the ten exchanges we first calculate all volatility measures separately for each exchange and then calculate the trading volume-weighted average across exchanges. For an exchange-currency combination to be included in the cross-sectional average we require that at least 90 daily return observation per calendar year are

[5] The only fiat currency traded on Poloniex is the Malaysian ringgit (RM). Because the USD-RM market is rather illiquid and the USD-RM exchange rate therefore potentially unreliable we decided to exclude from our sample currency pairs where a cryptocurrency is traded against RM on Poloniex. Consequently, only BTC prices from Poloniex are included in our sample.

[6] The exchange with the largest number of direct USD quotes is Bitfinex, with 42 direct quotes.

[7] Indirect prices are a possible cause for concern because it is known that there are arbitrage opportunities in cryptocurrency markets (e.g. Makarov and Schoar, 2020). We note, though, that the trading venues in our sample belong to the most liquid market places for cryptocurrencies, and higher liquidity tends to be associated with higher price efficiency (e.g. Wei, 2018). Using pairs for which both, direct and indirect prices, are available, we find only very small pricing errors.

available. We implement this procedure separately for two distinct sample periods, 2020 and 2021. However, we only include cryptocurrencies in our sample for which data for both years is available.

Table 2.1 shows descriptive statistics for the four samples, i.e. each combination of our BTC and USD sample with the sample periods 2020 and 2021. For each subsample the table provides information on the mean return and the two volatility measures. The most important insight from the descriptive statistics is that the volatility of the cryptocurrencies in our sample varies considerably in the cross-section. It is this variation that we wish to explain in our empirical analysis. The table further reveals that the volatility is generally higher in the BTC sample than in the USD sample.

In our empirical setting we have a limited number of cross-sectional observations (i.e. the cryptocurrencies) and a large number of potentially relevant explanatory variables (i.e. the design features). In a first step we reduce the number of explanatory variables by conflating some of the design feature variables.[8] Furthermore, we do not include the specific hash function employed by a cryptocurrency into our regression analysis but rather capture the effect on volatility of the hash function by its age.

The majority of our independent variables are binary variables. We code them such that the default value corresponds to the design of the Bitcoin network. Furthermore, we redefine continuous variables such that the value for Bitcoin is zero. For example, we recalculate blocktimes as

$$BlockTime_{mod} = \frac{Blocktime_{Bitcoin} - Blocktime}{Blocktime_{Bitcoin}}. \tag{2.1}$$

We proceed in a similar way for the age of the hash function. Given this definition of our independent variables all of them are zero for the Bitcoin network.

We use the machine-learning-based LASSO (absolute shrinkage and selection operator) procedure to select those variables that affect cryptocurrency volatility. The LASSO regression connects variable selection and regularization with 10-fold cross validation, which is repeated $10,000$ times in our analysis.[9]

2.3 Results

Tables 2.2 and 2.3 show the results for the two volatility measures. Each table shows separate results for the four subsamples (i.e. the intersection of our BTC and USD sample with the sample periods 2020 and 2021). If a variable is never selected by the LASSO procedure the

[8] We conflate the variables *CodeGo* and *CodeOther* to a single binary variable *CodeNonC++* which is set to one if at least one of the former variables is one, and zero otherwise. Similarly, we introduce the binary variables *CodeNonECDSA*, *IntentionNonPayment* and *Inflationary*. We refer the reader to Chapter 1 for further details.

[9] When we use five folds instead of ten in the cross-validation procedure our results remain qualitatively similar.

respective cell in the table has no entry. For all variables which are selected at least once we provide an estimate of the sign and strength of its impact on volatility. We further report how frequently a variable has been selected by the LASSO procedure. Specifically, *** [**, *, #] indicates that the respective variable has been selected in more than 80% [60%, 40%, 20%] of the cases.

When we measure volatility by the standard deviation of returns (Table 2.2) we find for the 2020 subsample that older cryptocurrencies are less volatile, and that volatility is higher for cryptocurrencies that were developed by private (i.e. for-profit) developers, for cryptocurrencies that use Proof-of-Stake or delegated Proof-of-Stake as consensus mechanism, and for cryptocurrencies that are developed for purposes other than being a means of payment. All these results are consistent in the BTC and USD subsamples. In contrast, these design features have essentially no explanatory power for volatility in the 2021 subsample. In this sample we only find (weak) evidence that forks are more volatile.

The interquartile range is a robust measure of volatility because it is not affected by outliers. As shown in Table 2.3 it delivers more significant results than the standard deviation of daily returns. These results confirm (for both the 2020 and 2021 samples) the findings that older cryptocurrencies are less volatile, and that cryptocurrencies developed by private for-profit entities and those developed for purposes other than being a means of payment are more volatile. The result that cryptocurrencies are more volatile that use Proof-of-Stake or delegated Proof-of-Stake as consensus mechanism is confirmed for the BTC sample but not for the USD sample.[10] Additionally, we find that cryptocurrencies with an obligation to pay a transaction fee and cryptocurrencies that use a cryptographic signature algorithm other than the Elliptic Curve Digital Signature Algorithm (ECDSA) are more volatile. The impact of other design features on volatility is weaker and/or is not consistent across the subsamples.

2.4 Conclusion

In this paper we analyze whether design features of cryptocurrencies affect their volatility. We report results for four subsamples (returns based on either BTC prices or USD prices, sample periods 2020 and 2021). We further report results for two volatility measures, the standard deviation and the interquartile range of returns. We put more emphasis on the results obtained using the interquartile range because this measure of volatility is robust to the presence of outliers in the return data. We find that older cryptocurrencies tend to be less volatile while currencies developed by private developer teams and those developed for purposes other than being a means of payment tend to be more volatile. Details of the cryptographic procedures and an obligation to pay transaction fees also appear to affect volatility, at least when measured by the interquartile range.

[10]This discrepancy between the results for the USD and the BTC samples is not caused by the fact that Bitcoin is included in the former but not in the latter. As noted previously, when we exclude Bitcoin from the USD sample we obtain results similar to those presented in the paper.

Table 2.1: Return Data: Descriptive Statistics

	Mean	First Quartile	Median	Third Quartile	Cross-sectional Standard Deviation
Panel A: *Measures of Returns in 2020 of Price in BTC*					
Mean Return	-0.0005	-0.0018	-0.0007	0.001	0.0029
Standard Deviation	0.0559	0.0417	0.0528	0.0613	0.0209
Interquartile Range	0.0485	0.0371	0.0471	0.0566	0.0158
Panel B: *Measures of Returns in 2020 of Price in USD*					
Mean Return	0.0043	0.003	0.004	0.0056	0.0028
Standard Deviation	0.0679	0.0574	0.0635	0.0743	0.0169
Interquartile Range	0.0588	0.048	0.057	0.0665	0.0143
Panel C: *Measures of Returns in 2021 of Price in BTC*					
Mean Return	0.0044	0.002	0.0036	0.0058	0.0037
Standard Deviation	0.0737	0.0598	0.0675	0.0832	0.0244
Interquartile Range	0.058	0.0486	0.0604	0.0662	0.0131
Panel D: *Measures of Returns in 2021 of Price in USD*					
Mean Return	0.0063	0.0043	0.0057	0.0082	0.0032
Standard Deviation	0.0857	0.0755	0.0809	0.0945	0.0189
Interquartile Range	0.0769	0.0668	0.0787	0.0849	0.0124

The table shows descriptive statistics (mean, quartiles, cross-sectional standard deviation) of the mean return, the standard deviation and interquartile range of returns. Panels A and B (C and D) show the summary statistics for the 2020 (2021) sample period; Panels A and C (B and D) show the summary statistics for the BTC (USD) sample.

Table 2.2: Lasso Results with Standard Deviation as Dependent Variable

	2020		2021	
Variables	(1) BTC	(2) USD	(3) BTC	(4) USD
Constant	0.0509***	0.0636***	0.0709***	0.0834***
DaysAge	-0.0002#	-0.0036***	-	-
DeveloperNPO	-	-	-	-
DeveloperPrivate	0.0000	0.0009***	-	-
MajorityChanges	-	-	-	0.0000
CodePublic	-	-	-	-
CodeNonC	-	-	-	-
Fork	-	-	0.0000***	0.0000***
ConsensusPoSDPoS	0.0068***	0.0033***	-	-
ConsensusOther	-	-	-	-
HashAge	-	-	-	-
CurveNonECDSA	-	-	-	-
MaxSupply	-	-	-	-
Deflationary	-	-	-	-
FixedSupply	-	-	-	-
RewardCoinbase	-	-	-	-
RewardAlternative	-	-	-	-
BlockTimeAverage	-	-	-	-
TransactionFeeObligation	-	-	-	-
TipSpecialTreatment	-	-	-	-
NoFeeTipForMinerForger	-	-	-	-
IntentionNonPayment	0.0011***	0.0028***	-	-
SmartContractSupport	-	-	-	-
TokenUsageBeyondPayment	-	-	-	-
LedgerStyleOther	-	-	-	-
AccountingBalance	-	-	-	-
Anonymous	-	-	-	-
NonAnonymous	-	-	-	-
∅ Observations	57	58	57	58
∅ R^2	0.1436	0.1680	0.0000	0.0002

This table reports the average of the parameter estimate indicating the economic significance when applying LASSO with return standard deviation as dependent variable and design feature variables as independent variables. The four columns present results for the 2020 and 2021 sample periods and for the USD and BTC samples. #, *, **, and *** indicate that the percentage of cases in which the variable is selected by the LASSO procedure are at least 20%, 40%, 60% and 80%, respectively; if a cell shows a figure without a superscript the corresponding variable is selected by the LASSO procedure in less than 20% of the cases. If a cell shows a " - " the corresponding variable is never selected by the LASSO procedure.

Table 2.3: Lasso Results with Interquartile Range as Dependent Variable

	2020		2021	
Variables	(1) BTC	(2) USD	(3) BTC	(4) USD
Constant	0.0459***	0.055***	0.0533***	0.0753***
DaysAge	-0.0017***	-0.0066***	-0.0101***	-0.0067***
DeveloperNPO	-	-	-	-
DeveloperPrivate	0.0001	0.0005**	0.0001#	0.0003*
MajorityChanges	-0.0000	-0.0006	-	-
CodePublic	-0.0000	-0.0000	-0.0000	-
CodeNonC	-	-	-0.0000	-
Fork	-	-	0.0010*	0.0001
ConsensusPoSDPoS	0.0005***	-	0.0025***	-
ConsensusOther	-	-	-	0.0000
HashAge	-	-	0.0000	0.0001
CurveNonECDSA	0.0045***	0.0064***	-0.0000	0.0002*
MaxSupply	-0.0000	-	-0.0033***	-0.0001
Deflationary	-	-	0.0038***	0.0000
FixedSupply	-	-	0.0000	0.0001
RewardCoinbase	-	-	-0.0007*	-0.0000
RewardAlternative	0.0000	0.0001	0.0018***	0.0001
BlockTimeAverage	-	-	0.0028***	0.0000
TransactionFeeObligation	0.0000	0.004***	0.0018***	0.0000
TipSpecialTreatment	-	-	0.0003*	0.0000
NoFeeTipForMinerForger	-0.0000	-0.0002	-0.0011*	-0.0001
IntentionNonPayment	0.0035***	0.0022***	0.0039***	0.0025***
SmartContractSupport	-	-	0.0001	0.0000
TokenUsageBeyondPayment	-	-0.0000	-0.0000	0.0000
LedgerStyleOther	-0.0000	-0.0009	-0.0008*	-
AccountingBalance	-	-	-0.0000	-
Anonymous	-	-	0.0027***	0.0001
NonAnonymous	-	-0.0002	-0.0000	-0.0000
∅ Observations	57	58	57	58
∅ R^2	0.1538	0.3019	0.4462	0.1724

This table reports the average of the parameter estimate indicating the economic significance when applying LASSO with the interquartile range as dependent variable and design feature variables as independent variables. The four columns present results for the 2020 and 2021 sample periods and for the USD and BTC samples. #, *, **, and *** indicate that the percentage of cases in which the variable is selected by the LASSO procedure are at least 20%, 40%, 60% and 80%, respectively; if a cell shows a figure without a superscript the corresponding variable is selected by the LASSO procedure in less than 20% of the cases. If a cell shows a " - " the corresponding variable is never selected by the LASSO procedure.

Chapter 3

Bitcoin Blackout: Proof-of-Work and the Risks of Mining Centralization[1]

3.1 Introduction

Cryptocurrencies have seen a remarkable increase in popularity since their inception only a few years ago. The most prominent one, Bitcoin, was the first to operate without any central authority by relying on a distributed ledger, the blockchain. Transactions conducted on the blockchain are recorded and verified through the proof-of-work (PoW) consensus mechanism, which addresses one of the fundamental problems of decentralized virtual currencies without trusted authorities, the double-spending problem. The process is usually referred to as "mining". To verify the integrity of transactions, the participants of the mining, or the miners, compete by solving computationally intensive but otherwise meaningless mathematical puzzles. The first to find a solution appends the next block of transactions to the chain and receives newly-issued coins and any fees paid by users. However, the process is very energy-intensive, so miners tend to gravitate towards regions with cheap electricity, undermining the network's decentralization. As the cryptocurrency market continues to evolve, concerns regarding the resulting limits to decentralization and the ecological impact of mining increase.[2]

[1]This work is the result of collaboration with Dr. Stefan Scharnowski. As a group of co-authors, we thank Louis Bertucci (discussant), Imad Chehade (discussant), Alfred Lehar, Daniel Rabetti (discussant), Erik Theissen, Victoria Treßel (discussant), Stefan Voigt, Christian Westheide, and conference and seminar participants at the 11th International Conference of the Financial Engineering and Banking Society, the 38th Annual Meeting of the French Finance Association, the 26th Spring Meeting of Young Economists, the Economics of Financial Technology Conference Edinburgh 2022, the CryptoAssets and Digital Asset Investment Conference Rennes 2022, the 15th RGS Doctoral Conference in Economics, the 4th UWA Blockchain and Cryptocurrency Conference, the Cryptocurrency Research Conference 2021, and the University of Mannheim for helpful comments and suggestions. We gratefully acknowledge financial support from the German Science Foundation (DFG) under grants TH 724/7-1.

[2]Please refer to the Appendix 3.6.1 for details about mining.

We examine short-term risks associated with this geographical centralization of mining by exploiting an exogenous shock to the electricity supply in a relatively small region with heavy Bitcoin mining activity as a quasi-natural experiment. During a blackout lasting several days, we document a drop of about 24% in the total computing power of the Bitcoin network. Compared to a control group consisting of a cryptocurrency using a more energy-efficient consensus mechanism, we find that the fees paid to miners to have transactions included in the blockchain increase substantially while the number and value of transactions decrease. Even though we find a shift to off-chain trading activity as seen in increased exchange trading volume, we also observe a drop in secondary market quality: Exchange rate volatility increases substantially and liquidity deteriorates as transaction costs such as bid-ask spreads increase. Furthermore, we document that market integration diminishes as price differences between various cryptocurrency exchanges widen.

The results have important implications regarding the centralization within the Bitcoin network. Operational and geopolitical risks stemming from the consensus mechanism's dependence on cheap electricity have the potential to adversely impact the entire network. Since we document strong spillover effects to exchange trading activity, these risks potentially affect a wide range of market participants. Traders and regulators should be aware of these risks associated with energy-intensive proof-of-work networks.

We contribute to several streams within the literature. Most directly, we contribute to the stream analyzing centralization within cryptocurrency networks. Closely related to our paper, Makarov and Schoar (2022) estimate the geographical concentration of mining using a novel approach based on the location of the cryptocurrency exchange a miner uses to cash out their profits. Finding that most of mining is located in China, they also consider the same exogenous shock we study to verify their approach. However, they do not analyze any risks or market quality implications associated with this type of mining centralization.

While we focus on geographical concentration, economic forces might lead to centralization along several other dimensions. For example, Böhme et al. (2015) identify cryptocurrency exchanges and wallet services as potential sources of centralization. Mining pools where users combine their resources to obtain a more stable stream of mining rewards are also considered a threat to decentralization, though the model of Cong et al. (2021) expects a decentralized market structure for Bitcoin in the long run even in the presence of centralized mining pools because miners' cross-pool diversification and pool managers' endogenous fee setting makes larger pools to attract disproportionately fewer miners and hence grow more slowly. Focusing on another source of centralization, Arnosti and Weinberg (2022) find that production and ownership of mining hardware lead to a costly arms race and centralization within Bitcoin mining. Similarly, Ferreira et al. (2022) model blockchain governance and show that in a PoW system, so-called blockchain conglomerates – large firms that operate in multiple blockchain-related businesses such as mining equipment and mining pools – may control blockchain votes and thus governance. Capponi et al. (2021) argue that because miners are capacity-constrained, centralization does not necessarily result from increases in hardware efficiency. Instead, investments leading to more efficient mining hardware allow new and small miners to enter or expand their operations. Lehar and Parlour (2022) investigate the effect of strategic capacity management by miners and show that mining concentration is related to higher levels of fees for users. While they also find that fees paid to miners increase during the same event period, especially for impatient traders, they do not focus on the wider systemic implications of geographical mining centralization in PoW networks. In this paper, we add to the literature on the centralization within the Bitcoin network by explicitly showing multiple dimensions of system-wide risks arising from geographically concentrated mining.

Furthermore, no prior study empirically investigates differences in geographical concentration and associated risks across different consensus protocols. According to Irresberger et al. (2021), the most common consensus mechanism is PoW, followed by proof-of-stake (PoS) and its derivatives.[3] By comparing two currencies with different consensus mechanisms, our study also relates to the literature on the relative advantages of PoW and alternative mechanisms. In particular, Arnosti and Weinberg (2022) conjecture in their conclusion that under certain conditions, the PoS consensus mechanism might contribute to cryptocurrency decentralization. This mechanism does not rely on miners to verify transactions, but instead randomly chooses validators where the probability of being drawn increases in the amount of coins deposited as stake. Chosen validators update the blockchain and get rewarded with newly-issued coins. Their stake in the currency incentivizes validators not to compromise the blockchain, which would render the currency worthless. Within the PoS literature, Saleh (2021) provides a first economic analysis of the mechanism and gives equilibrium conditions for consensus. Still, a common criticism of the PoS mechanism is that it could lead to wealth accumulation and thus centralization because validators with larger stakes have a higher probability of being chosen and thus obtaining the reward, increasing their stake further. Roşu and Saleh (2021) address this concern by showing that under certain assumptions, any investor's share in the network follows a martingale and is not expected to change in the long run. Using simulations, Mueller-Bloch et al. (2022) investigate potential determinants of the level of decentralization within PoS networks. For example, having many potential validators and high transaction activity on the blockchain is associated with a more decentralized network.

An argument favoring non-PoW currencies is that they are generally more energy efficient. This not only contributes to geographical decentralization, but is also particularly important as several studies have voiced concerns that Bitcoin mining could be a substantial factor for

[3] The second largest cryptocurrency, Ethereum, transitioned from PoW to PoS in September 2022. We also note that other algorithms are sometimes used. Related to PoS is delegated proof-of-stake (DPoS), where stakeholders vote from a fixed and limited number of delegates. The voting power is proportional to the stake in the network and the voted delegate then is responsible for producing blocks. This mechanism is e.g. used by the cryptocurrencies Tronix (of the TRON network) and EOS. Leased proof-of-stake (LPoS) is similar to DPoS, but instead of voting, users lease out their tokens and thus the right to produce blocks to block-generating nodes in exchange for some award. LPoS does not rely on a limited number of nodes and thus tends to be less centralized than DPoS. Waves is the most prominent example of an LPoS cryptocurrency. Some alternative consensus mechanisms rely on storage space. For example, in Burstcoin's proof-of-capacity (PoC), miners compete by providing disk space that is not otherwise used and in Filecoin's proof-of-spacetime consensus mechanism, miners effectively rent out disk space. Other mechanisms include proof-of-importance, proof-of-burn, proof-of-elapsed-time, proof-of-authority, and the more general proof-of-weight. Some cryptocurrencies use combinations of different consensus algorithms.

climate change, though there is no consensus on the magnitude (Dittmar and Praktiknjo, 2019; de Vries, 2020; de Vries et al., 2022). Additionally, the economic significance of mining activities' spillover effects on the electricity market is a noteworthy concern for regulators and market participants. For the United States, Benetton et al. (2022) document higher electricity costs for households and small businesses due to regional cryptocurrency mining activity. For China, the authors find a crowding-out effect on the local economy due to mining-induced regional electricity rationing. Karmakar et al. (2021) observe that Bitcoin mining activity is related to increasing volatility levels in the electricity spot market. In a similar vein, Corbet et al. (2021) find a positive relationship between Bitcoin returns and the volatility of returns of certain electricity and utility companies. More generally, several studies have documented interdependencies and spillovers between cryptocurrency markets and energy markets (Ji et al., 2019; Pham et al., 2022; Ren and Lucey, 2022; Sharif et al., 2023). Since we compare Bitcoin and its energy-intensive consensus mechanism to a more energy-efficient alternative during a shock to the electricity market, our study also relates to this stream within the literature.

A noteworthy point about our analysis is that the shock to the network is only temporary. While in our case, this is a result of the limited duration of the blackout, the temporary nature of shocks to mining activity is also an inherent feature of the PoW mechanism. The difficulty of the mathematical puzzles miners have to solve is adjusted according to the total computing power of the network, albeit with a lag. For Bitcoin, the adjustment happens roughly every two weeks to keep the expected time between blocks at 10 minutes. This delayed but automatic adjustment ensures that any shock only temporarily affects the speed of settlement, attenuating the geopolitical and operational risks of centralization in the long term. While the difficulty of the aforementioned mathematical puzzles was adjusted downwards due to the blackout, the difficulty is adjusted upwards after the power was restored and hence total computing power of the network. In other words, the change in the difficulty was only temporary. In contrast to the blackout, the escalating investments by Bitcoin miners in mining equipment result in a higher hashrate and, consequently a temporarily reduced average block time.[4] However, over a longer horizon, this arms race does not significantly affect the Bitcoin network; instead, it prompts increases in difficulty to sustain an average block time of approximately ten minutes. In particular, the total rewards in the form of newly mined bitcoins stay the same (Capponi et al., 2021). However, in the short term, there can be substantial deviations from this long-run average settlement speed and the corresponding capacity of the blockchain.

In this respect, our study is thus in the spirit of the literature on market quality breakdowns and flash crashes: While these are temporary and of even shorter duration than the blackout, they have sparked the interest of academics and regulators since they may still be extremely costly to market participants (Gao and Mizrach, 2016; Kirilenko et al., 2017; Menkveld and Yueshen, 2019).

A related aspect to the limited duration of the shock to the network is often referred to

[4]Due to the unavailability of the exact time of power restoration, we are unable to incorporate the restoration of power as an additional shock in our study.

as network resiliency. The corresponding literature is typically concerned with the security of a blockchain in the long run, for example with respect to double-spending attempts. In this stream, Bertucci et al. (2022) model the Bitcoin mining industry and show that the security of the Bitcoin network is the result of a long-run equilibrium of mining revenues. However, in the medium run, shocks can have a substantial effect since the supply of mining hardware can only slowly adjust. John et al. (2021) compare PoW and PoS blockchains with respect to their scalability. Blockchains that scale better with the demand for transaction recording exhibit less congestion and lower fees. Importantly, the associated reduction in PoW miners' revenues lowers their incentives to invest in computational power, thus lowering the security of the blockchain. The opposite effect on security is found for PoS blockchains. In parallel work to ours, Stinner and Tyrell (2022) consider the resiliency of the Bitcoin network by exploiting two shocks to mining activity. However, their focus is on the security of the network, the incentive compatibility of mining even in extreme conditions, and how traders value the blockchain's stability. Conversely, our focus is on the systemic implications of mining centralization on blockchain activity and secondary market quality.

The remainder of the paper is structured as follows: Section 3.2 introduces the event and develops our hypotheses while Section 3.3 explains the data and the empirical methodology. Section 3.4 discusses the results for blockchain activity and secondary market quality before Section 3.5 concludes.

3.2 Background and Hypotheses

3.2.1 Bitcoin Mining and the Blackout

Cryptocurrency miners tend to be secretive regarding their operations, including their precise geographic location. However, because electricity constitutes the main input factor for miners of PoW currencies, they gravitate towards countries with cheap energy. During our sample period, this is in particular China (Delgado-Mohatar et al., 2019). While during the wet season from May to October, many miners are located near hydroelectric plants in the provinces of Sichuan and Yunnan, operations migrate to other areas in the dry season, particularly Xinjiang with its cheap coal energy. According to the Cambridge Centre for Alternative Finance (2021), China accounts for about 46% of all Bitcoin mining during our sample period. Moreover, Xinjiang alone accounts for about 25% of worldwide activity.

On April 10, 2021, a coal mine flooded in Hutubi County in northern Xinjiang, trapping 21 people. As a consequence, the local government announced extensive safety inspections on April 15, which led to the temporary shutdown of several power plants in the region and resulted in a local electricity shortage (Altxw.com, 2021). Various news articles then document a sharp decrease in the Bitcoin hashrate, i.e., the total computing power of the network, starting April 16 as cryptocurrency mining operations were deprived of electricity (e.g. Fortune.com, 2021; Digiconomist.net, 2021). Power only gradually resumed after about one week, leading to a gradual increase of the hashrate back to its previous level (Theblockcrypto.com, 2021). We hence consider the window from April 16 to April 22 as the blackout period. While we can reasonably precisely timestamp the beginning of the blackout, there is some uncertainty regarding when exactly power was restored. We address this issue in two ways: First, we estimate the implied hashrate and test not only if there is a significant drop during this period, but also if the computing power is restored back to its previous level afterwards. Second, we repeat our regression analysis using different window lengths and find that our results are robust to different specifications.

The blackout provides a unique opportunity to analyze risks associated with the geographical concentration of Bitcoin mining. Firstly, the blackout can be relatively accurately timestamped and immediately affected miners in that region. This is not necessarily the case for other types of events. For example, while government restrictions on cryptocurrencies and mining in particular are highly relevant signals regarding the prospects of cryptocurrencies, they are unlikely to immediately and effectively shut down all local mining operations, making a clear identification difficult (see e.g. Chen and Liu, 2022). Furthermore, the vulnerability of cryptocurrency mining to government restrictions is a consequence of geographical concentration stemming from the PoW consensus mechanism's energy dependence. Investigating the effects of the blackout is thus a more direct test of geographical mining concentration. Secondly, the considered blackout constitutes an exogenous shock to mining, which might not hold for other events. For instance, the electricity consumption of Bitcoin mining and the resulting stress to the power grid has been blamed for power outages in other regions like Iran (CNBC, 2021). An alternative perspective is that the blackouts occurred independently of Bitcoin mining, making those blackouts shocks to mining. Finally, many alternative shocks only indirectly affect mining, for example through cryptocurrency price changes. In contrast, our event directly impacts not only the profitability, but also the possibility of mining in the affected region and hence the mining capacity of the whole network, which is a direct consequence of the geographical centralization of mining. The consequence of geographical centralization is also illustrated by the 15% drop of hashrate after Kazakhstan shut down country-wide internet access in response to protests in January 2022 (CNBC, 2022).[5]

3.2.2 Hypothesis Development

If mining operations are indeed interrupted by the blackout, we expect the total computing power of the network to decrease. We hence first confirm that the hashrate of the Bitcoin

[5]In July and September 2021, China banned mining activity and cryptocurrency transactions. The exact impact of these bans is still debated. For example, according to the Cambridge Centre for Alternative Finance (2021) Bitcoin Mining Map, in August 2021 there was no Bitcoin mining in China. However, in September 2021 and January 2022, the estimate is at about 22% and 21% of global mining activity, respectively. A potential problem with this data is that it relies on the IP addresses of mining facilities, which after the ban may have been disguised using VPN tunnels and other measures. But even if many miners should have subsequently left Xinjiang and the rest of China, mining activity likely migrated to places with cheap electricity such as Kazakhstan and the United States, in particular Texas (de Vries et al., 2022). Geographical centralization of mining is hence still a potential problem for the whole network, even though miners may no longer specifically depend on cheap coal energy from Xinjiang.

network is significantly lower during the blackout. Finding evidence in favor of this first hypothesis thus validates using the blackout as a quasi-natural experiment.

Hypothesis 1: The hashrate of the Bitcoin network is lower during the blackout.

Keeping the difficulty of mining a block constant, a reduction in the hashrate leads to an increase in the average time between blocks and fewer mined blocks overall. The capacity on the blockchain is thus reduced and becomes more binding. Consequently, impatient traders compete for the scarce resource of blockchain capacity by bidding up fees, akin to a congestion of the blockchain as in Kim (2020) and Sokolov (2021). This reasoning also aligns with Easley et al. (2019). In their model, fees reflect the queuing problem faced by Bitcoin users. As waiting time increases, some users choose to increase their fees. However, others exit and do not submit transactions as the benefit of having the transaction recorded does not outweigh the cost associated with fees and waiting times. Strategic capacity management by miners to increase fee revenues as in Lehar and Parlour (2022) would further amplify this increase in fees even if the blockchain is running below its true capacity.

Hypothesis 2: Fees paid to miners increase during the blackout.

Reacting to the higher level of fees and the longer block time caused by the lower overall capacity of the blockchain, patient traders or those with only marginal utility gains from trade omit or postpone their trading as in Easley et al. (2019), leading to fewer transactions recorded on the blockchain.[6] While the effect is partially mechanical due to the increase in blocktime caused by capacity constraint, it is also related to the strategic behavior of miners, aligning with hypothesis 2. Given that many traders are willing to pay higher fees for high-value transactions, by selectively confirming transactions with high fees, the miners are also selectively confirming transaction with high value, resulting in higher average value of transaction recorded on the Bitcoin blockchain during the blackout. In other words, the decrease in the value of transactions recorded on the Bitcoin blockchain is not only mechanically caused by the reduction in the number of transaction recorded on the Bitcoin blockchain.

Hypothesis 3a: The number and value of transactions recorded on the Bitcoin blockchain decrease during the blackout.

Hypothesis 3b: The average value of transaction recorded on the Bitcoin blockchain increases during the blackout.

[6] Some traders might still submit their transactions during the blackout, but may only be willing to pay fees that are too low to incentivize miners to include the transactions in new blocks. These transactions would remain unconfirmed by the network until fees become low enough, which is economically similar to postponing the transactions.

So far, we have only considered on-blockchain trading activity. However, we also expect the blackout to impact off-chain activity, i.e., cryptocurrency trading activity on centralized exchanges. In the model of Zimmerman (2020b), limited settlement space and the associated competition between cryptocurrency users and speculators leads to a crowding-out of those who want to use the currency for payments. This decreases the efficacy of the currency as a means of payment. Similarly, it is expected that the blackout, disadvantageous to users seeking to utilize the currency for payments due to more limited blockchain capacity, will lead to a decrease in the return of Bitcoin during the blackout period.

Hypothesis 4: Bitcoin depreciates during the blackout.

The model proposed by Zimmerman (2020b) further suggests that the displacement of users intending to use the currency for payments would elevate its risk profile in terms of price volatility. Analogously, the volatility of Bitcoin is expected to increase during the blackout.

Hypothesis 5: Bitcoin returns become more volatile during the blackout.

The reasoning that leads to hypothesis 4 and 5 are also consistent with the equilibrium model of Biais et al. (2023), where the fundamental value of a cryptocurrency is derived from future transactional benefits. Furthermore, Bhambhwani et al. (2021) empirically show that there is a positive relationship between cryptocurrency prices and computing power. They conjecture that computing power is a fundamental pricing factor that proxies for systemic risk.

The increase in risk and the general uncertainty surrounding the blackout likely lead to a reduction in liquidity as market makers reduce their exposure. Furthermore, increases in settlement times may make it more difficult for market makers to manage their inventories.

Hypothesis 6: Secondary market liquidity deteriorates during the blackout.

Higher volatility and lower liquidity make it more difficult for arbitrageurs to exploit arbitrage opportunities, for example across different trading venues (Shleifer and Vishny, 1997; Roll et al., 2007). Furthermore, according to the model of Hautsch et al. (2021), settlement latency reduces cross-market trading and makes exploiting potential arbitrage opportunities riskier. This is because arbitrageurs on centralized exchanges face the dilemma of either trusting the possibly unreliable exchanges by keeping funds in their accounts or potentially missing out on trading opportunities due to high blockchain latency which reduces the speed of moving funds into and out of the exchange accounts used for trading. Since centralized cryptocurrency exchanges have a long history of misplacing customer funds by being hacked or by embezzlement, trust in many exchanges is generally relatively low. Hence, most traders prefer not to keep large amounts of funds with an exchange for longer periods of time (Hoang and Baur, 2022). This leads to larger and more volatile cross-venue price differences when settlement latency is high.

Hypothesis 7: Secondary market integration decreases as cross-venue price differences increase.

3.3 Empirical Approach

3.3.1 Sample Selection and Data

The sample period includes the blackout period from April 16 to April 22 and one week before, though we additionally obtain data for the surrounding weeks to analyze any trends. During this time, we compare Bitcoin to Ada, the internal cryptocurrency of the Cardano platform. Cardano was launched in 2017 and consists of two layers, where the first layer is the settlement layer tracking Ada ownership, similar to the Bitcoin network. The second layer facilitates smart contracts akin to the Ethereum network. Importantly, Cardano uses the PoS consensus mechanism, making Ada the largest PoS and non-PoW cryptocurrency at the time of the blackout, representing more than half of the market capitalization of all PoS currencies (see also Irresberger et al., 2021). Overall, it is the fifth-largest cryptocurrency as of May 2021. Therefore, Ada can be regarded as the equivalent of Bitcoin among non-PoW crytocurrecies, but, contrary to Bitcoin, is notably less impacted by the blackout due to significantly lower energy consumption for transaction verification. This make Ada suitable as the control currency. However, to address concerns regarding our choice of control currency, we additionally obtain data on three other non-PoW cryptocurrencies and use these for robustness tests.

Our data comes from two main sources: First, we collect data from the currencies' blockchains: the number of new blocks, the average time between two blocks, and the number of transactions contained in all blocks. Furthermore, we compute the value sent within all transactions, the average value of a transaction, and the total fees paid to miners using market prices. We then compute the value-weighted average relative fee as given by total fees paid to miners within a block divided by the value of all transactions within that block. However, these average fees per block may mask substantial heterogeneity in the fees paid for individual transactions, for example due to strategic capacity management by miners as in Lehar and Parlour (2022). We hence also look at the within-block distribution of fees and compute the 10th and 90th percentile of equally-weighted relative fees within a block as well as their standard deviation.

Second, we download minutely exchange rate data from Kraken, which is generally considered trustworthy and one of the most liquid exchanges for trading Bitcoin and Ada. We focus on trading cryptocurrencies against the US dollar. For robustness and to study cross-venue market integration, we also consider Coinbase and Binance. At the latter, Bitcoin and Ada are traded against the stablecoin Tether.[7] According to Makarov and Schoar (2022), Binance

[7]While there is a debate on the stability of stablecoins (Hoang and Baur, 2021), differences between

is one of the most popular exchanges for Chinese miners to cash out their accumulated mining rewards.

We aggregate the exchange data to an hourly frequency by calculating logarithmic returns based on hourly closing prices, the standard deviation of minutely logarithmic returns, and the trading volume in USD. Because Scharnowski (2021a) indicates that Bitcoin liquidity is related to the hashrate, we decide to also study the change of Bitcoin liquidity caused by the event. We follow Brauneis et al. (2021) and consider different measures of liquidity. Firstly, we use the Corwin and Schultz (2012) high-low spread estimator, which according to Brauneis et al. (2021) performs well in capturing the time-series variation of liquidity.

$$\text{Spread}_t = \frac{2\left(\exp\left(\alpha\right)-1\right)}{1+\exp\left(\alpha\right)} \quad \text{where} \quad \alpha = \frac{\sqrt{2\beta}-\sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}$$

$$\beta = \left[\ln\left(\frac{\text{High}_t}{\text{Low}_t}\right)\right]^2 + \left[\ln\left(\frac{\text{High}_{t+1}}{\text{Low}_{t+1}}\right)\right]^2 \qquad \gamma = \left[\ln\left(\frac{\max\left(\text{High}_t, \text{High}_{t+1}\right)}{\min\left(\text{Low}_t, \text{Low}_{t+1}\right)}\right)\right]^2$$

As commonly done, we set negative spread estimates to zero.

Secondly, we compute the Kyle and Obizhaeva (2016) illiquidity index, which performs well in capturing the cross-sectional variation of liquidity. Intuitively, the measure expresses how volatile returns react to a given trading volume. Formally, we use the following specification

$$\text{Illiquidity}_t = \left[\frac{\sigma_t^2}{\text{Volume}_t}\right]^{\frac{1}{3}}$$

Finally, to study any changes in market integration, we compute price differences between various trading venues based on minutely closing prices.

$$\text{Price Diff}_t = \left|\ln\left(\frac{\text{Close}_{\text{Exchange A},t}}{\text{Close}_{\text{Exchange B},t}}\right)\right|$$

We consider the average, the standard deviation, and the 90th percentile of price differences during each hourly interval to capture different dimensions of the distribution of cross-venue price differences.

Table 3.1 provides summary statistics based on two weeks before the blackout.[8] While both currencies are the largest by market capitalization among the cryptocurrencies using their respective consensus mechanism, Bitcoin's transaction activity is still much higher. To reduce noise and bring the metrics to comparable levels, we hence use data from that week to standardize the variables. In particular, for each currency and for all variables except returns, we subtract the time-series mean and divide by the standard deviation.

USD and Tether are small during our sample period. The maximum and minimum of the USDT/USD exchange rate at Kraken is 1.0061 and 0.9990, respectively. The choice of quote currency is thus unlikely to meaningfully impact our results. Still, in most analyses we only compare different base currencies while keeping the quote currency fixed.

[8]Summary statistics based on data during the blackout period can be found in the appendix.

Table 3.1: Descriptive Statistics

	Mean	SD	P5	P50	P95	Skew.	Kurt.
Panel A: *Bitcoin*							
$\text{Transactions}_{\text{N}}$	12.23	4.84	5.23	12.00	20.74	0.4	2.8
$\text{Transactions}_{\text{Total Value}}$	3.83	1.99	1.13	3.43	7.66	1.0	4.2
$\text{Transactions}_{\text{Size}}$	0.33	0.17	0.12	0.31	0.61	1.9	11.2
$\text{Transactions}_{\text{Block}}$	2068.46	425.19	1241.33	2094.95	2683.20	−0.3	2.8
Blocktime	11.72	6.38	5.61	10.01	22.12	2.2	9.8
$\text{Fees}_{\text{Total}}$	221.84	85.53	109.27	205.02	380.06	0.6	3.1
$\text{Fees}_{\text{Relative}}$	0.69	0.32	0.34	0.63	1.26	1.4	5.6
$\text{Fees}_{\text{SD within Block}}$	34.18	13.89	19.28	31.82	54.58	3.2	22.5
$\text{Fees}_{\text{P10 within Block}}$	0.02	0.01	0.01	0.02	0.03	0.7	3.6
$\text{Fees}_{\text{P90 within Block}}$	10.83	2.61	7.21	10.74	14.45	0.9	6.4
Return	−0.71	45.04	−87.36	−3.00	68.89	0.5	4.9
Volatility	6.29	2.67	3.36	5.68	11.52	1.5	5.7
Volume	8.15	6.69	2.07	6.52	20.11	3.0	17.9
Spread	0.50	0.42	0.08	0.36	1.35	1.8	6.7
Illiquidity	1.37	0.32	0.91	1.37	1.94	0.2	2.4
Cross-venue Price $\text{Diff.}_{\text{Mean}}$	3.71	1.27	2.31	3.44	5.72	2.4	13.2
Cross-venue Price $\text{Diff.}_{\text{P90}}$	7.54	2.10	5.11	7.16	10.92	1.8	9.7
Cross-venue Price $\text{Diff.}_{\text{Std.Dev.}}$	2.69	0.72	1.88	2.56	3.67	3.0	19.5
Panel B: *Ada*							
$\text{Transactions}_{\text{N}}$	1.49	0.72	0.94	1.41	2.16	8.1	87.4
$\text{Transactions}_{\text{Total Value}}$	0.20	0.18	0.08	0.16	0.40	5.0	35.5
$\text{Transactions}_{\text{Size}}$	0.14	0.11	0.06	0.11	0.28	4.2	25.6
$\text{Transactions}_{\text{Block}}$	9.49	4.62	6.18	8.94	13.57	8.5	94.7
Blocktime	0.38	0.03	0.34	0.38	0.44	0.6	3.5
$\text{Fees}_{\text{Total}}$	0.39	0.17	0.24	0.37	0.60	5.7	52.3
$\text{Fees}_{\text{Relative}}$	0.03	0.01	0.01	0.02	0.04	2.1	13.4
$\text{Fees}_{\text{SD within Block}}$	0.84	0.27	0.49	0.82	1.41	0.8	4.3
$\text{Fees}_{\text{P10 within Block}}$	0.05	0.09	0.01	0.03	0.13	5.2	33.9
$\text{Fees}_{\text{P90 within Block}}$	1.21	0.47	0.63	1.06	2.09	0.9	3.5
Return	1.70	92.54	−137.86	4.72	139.98	2.5	22.3
Volatility	11.46	8.04	4.33	9.43	23.87	3.4	20.0
Volume	1.16	1.45	0.29	0.82	2.67	6.2	52.9
Spread	0.78	1.08	0.09	0.53	2.21	6.8	64.3
Illiquidity	6.47	20.06	2.88	4.84	7.52	12.8	164.4
Cross-venue Price $\text{Diff.}_{\text{Mean}}$	6.25	1.77	4.05	5.96	9.40	1.2	5.0
Cross-venue Price $\text{Diff.}_{\text{P90}}$	12.48	3.15	8.44	11.91	18.26	0.7	3.0
Cross-venue Price $\text{Diff.}_{\text{Std.Dev.}}$	4.63	1.25	3.08	4.45	6.93	1.3	6.0

This table shows summary statistics based on hourly data from the week of April 2. $Transactions_{N}$ is the number of transactions recorded on the blockchain in 1k, $Transactions_{Total\ Value}$ their value in USD 1bn, and $Transactions_{Size}$ their average size in USD 1mn. $Transactions_{Block}$ gives the average number of transactions per block and *Blocktime* the average time between two blocks in minutes. $Fees_{Total}$ is the sum of all fees paid by users in USD 1k and $Fees_{Relative}$ the same relative to the value of the transactions in basis points. $Fees_{P10/P90\ within\ Block}$ is the hourly average of the 10th/90th percentile of equally-weighted relative fees within a block in percentage points and $Fees_{SD\ within\ Block}$ the hourly average of the standard deviation of equally-weighted relative fees within a block in percentage points. The remaining variables are based on trading data from Kraken: *Return* is the logarithmic return of hourly closing prices in basis points, *Volatility* the standard deviation of minutely log returns in basis points, *Volume* is the trading volume in USD 1mn, *Spread* is the high-low spread estimate in basis points, and *Illiquidity* is the illiquidity index by Kyle and Obizhaeva (2016) in basis points. *Cross-venue Price* $Diff._{Mean/P90/Std.Dev.}$ is the hourly average/90th percentile/standard deviation of relative USD price difference between the exchanges of Kraken and Binance in basis points.

3.3.2 Estimating the Network Hashrate

We first document the extent to which Bitcoin miners were affected by the blackout. While not directly observable, their aggregate computing power can be estimated by the implied hashrate. The hashrate is related to the ratio of the current difficulty within the network and the time between two blocks. The former describes how computationally demanding it is to solve the mathematical puzzle by finding a hash below a certain threshold which would lead to the successful mining of a block. Importantly, while the time between blocks is stochastic, the difficulty is adjusted roughly every two weeks to keep the expected time between blocks at ten minutes. Importantly, the difficulty stayed constant during our sample period. Several days later there was a sharp drop in difficulty, reflecting the lower average hashrate during the blackout.

Using the observed average time between two blocks during some interval then gives an estimate of the implied network hashrate. Regarding the choice of time interval, there is a trade-off between the stability of the estimate and the frequency with which we can observe the computational power of the network. We use an interval length of three hours, but our results are generally robust to using longer and shorter intervals.[9]

$$\widehat{\text{Hashrate}}_t = 2^{32} \times \frac{\text{Difficulty}_t}{\text{TimeBetweenBlocks}_t}$$

3.3.3 Regression Analysis

We analyze the impact of the blackout in a difference-in-difference framework by comparing Bitcoin to a non-PoW cryptocurrency that, by design, only trivially depends on electricity and is thus not directly affected by the blackout. Using this approach, we control for unobservable confounding factors that affect both cryptocurrencies, for example through economy-wide or (crypto)market-wide changes, while isolating the effect on Bitcoin with its PoW consensus mechanism.[10]

Specifically, for the period including the blackout and the one week before, we estimate

$$\text{Y}_{it} = \alpha + \beta_1 \text{Bitcoin}_i + \beta_2 \text{Blackout}_t + \beta_3 \text{Bitcoin}_i \times \text{Blackout}_t + \varepsilon_{it}$$

where Y_{it} is a standardized measure of blockchain or trading activity of cryptocurrency i at time t and *Bitcoin* and *Blackout* are binary indicator variables. The constant shows the value

[9]Note that the subscript of t is needed for the varaible "Difficulty" as it is adjusted roughly every two weeks, as mentioned in Section 3.1, although it is constant during our sample period.

[10]While in this study we do not consider other PoW currencies, it is ultimately an empirical question whether miners of other PoW cryptocurrencies were located in the same region affected by the blackout. However, using data from etherscan.io, we find that the estimated Ethereum network hashrate was not impacted by the blackout, increasing marginally by about 0.22% during the blackout period compared to the week before. This might be caused by the fact that Ethereum mining is less based in China than Bitcoin mining, which is consistent with the observation by Neumüller (2023). Hence, we empirically justify the inclusion of only Bitcoin in the PoW treatment group for our analysis, as opposed to including alternatives such as Ethereum.

of Ada before the blackout, normalized using data from two weeks before. The coefficient for Bitcoin gives the normalized difference between the two currencies before the blackout. The blackout coefficient shows the normalized effect of the blackout, while the interaction term gives its additional effect on Bitcoin. We estimate the equation using ordinary least squares and use heteroskedasticity-robust standard errors.[11]

A crucial assumption of our difference-in-difference estimator is that the two cryptocurrencies exhibit parallel trends before the event. To verify the validity of this assumption, we plot the development of the different variables for both currencies. Before the blackout, differences are generally small and the two currencies closely co-move. We also test the assumption more formally by comparing the average changes in the hourly variables between the two currencies during the two weeks before the blackout. The (untabulated) differences are insignificant in all cases, again suggesting that the parallel trends assumption indeed holds.

Another concern might be that there are spillover effects from Bitcoin as the leading cryptocurrency to the rest of the market, which then would indirectly affect the control currency. However, assuming such spillovers impact the control currency in the same direction, they would actually lead to an underestimation of the additional effect the blackout has on Bitcoin and would instead be captured by the coefficient for the blackout.[12] Moreover, capturing this spillover effect and thus the impact of the blackout on the wider cryptocurrency market in the coefficient for the Blackout allows us to draw conclusions regarding the propagation of the risks inherent in Bitcoin's geographical centralization to other cryptocurrencies as well, thus showing the systemic risks within the cryptocurrency market arising from one currencies' operational risks due to geographical concentration.

3.4 Results

3.4.1 Drop in Hashrate

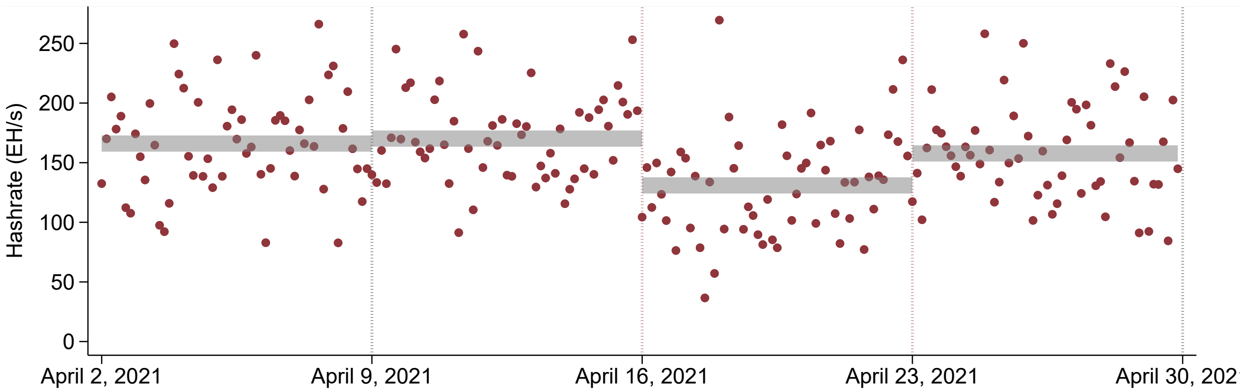
Figure 3.1 shows the development of the implied hashrate of the Bitcoin network. While volatile, the average during the two weeks before the blackout is virtually identical at about 170 Exahashes per second (Eh/s).[13] During the blackout, the hashrate drops to 130 Eh/s, or by about 24%, suggesting that about one quarter of worldwide Bitcoin mining operations were affected by the blackout. This closely aligns with the estimate of the Cambridge Centre for Alternative Finance (2021) according to which 25% of worldwide mining activity is located in Xinjiang during our sample period, suggesting that even within this region, all mining operations are located in close geographical proximity. The drop is highly statistically significant: t-tests for differences between the blackout period and the week before

[11] Our results do not depend on the exact specification of standard errors. Non-robust standard errors yield virtually identical results while clustering by the date gives statistically stronger results.

[12] This underestimation becomes apparent later when we consider abnormal returns to calculate volatility.

[13] One exahash per second corresponds to one quintillion (10^{18}) hashes per seconds while one hash essentially corresponds to one guess for the mathematical puzzle the miners try to solve.

Figure 3.1: Implied Hashrate of the Bitcoin Network



This graph shows the estimated implied hashrate of the Bitcoin network in exahash per second. Each estimate is based on the current difficulty of the network and the average time between blocks mined within a three-hour window. The vertical dashed lines indicate the beginning and end of the blackout and the three surrounding weeks. The horizontal bars show the average hashrate during the respective windows.

or afterwards give absolute test statistics of 5.19 and 3.37, respectively. Using the median instead of the average time between two blocks to estimate the hashrate gives similar results, though at an overall higher level. The difference between the implied hashrate during the week before to the week after the blackout is statistically insignificant, supporting the choice of the length of the blackout window.

To further validate the estimation of the hashrate, we then ask whether the magnitude of the drop is plausible given the typical capacity of a coal power plant. Assuming a mining efficiency of 20 $\frac{\text{Th/s}}{\text{kW}}$, the drop of about 40 Eh/s corresponds to a decrease in electrical power drawn by miners of 2,000 MW.[14] While the exact power plant(s) affected by the blackout are unknown, this is about the same order of magnitude as the capacity of a single large coal power plant. For example, the Wujiaqu power station, also in Xinjiang, has a total capacity of 3,640 MW, though most power plants in the region are likely smaller. The observed drop in implied hashrate is thus consistent with the shutdown of a single larger or several smaller coal power plants.

The results for the implied hashrate can also be expressed as the time between two blocks. During the blackout, the average blocktime is 16.37 minutes. Compared to the long-term average of 10 minutes and given the rule of thumb of waiting for six blocks to consider a transaction confirmed, Bitcoin traders have to wait for an additional 38 minutes – or 64% longer than usual – to be sure that their transactions are irreversibly recorded on the blockchain. There is hence a both statistically and economically significant increase in settlement latency.

[14]Hashrate is measured in units of hash/second, or how many hashing calculations per second can be performed. Th/s means one trillion hashes per second; Eh/s means one quintillion hashes per second. bitFlyer EUROPE S.A. (2019). Our assumption of mining efficiency is based on the mining equipment available at the time. For example, the Bitmain Antminer S15 has a hashrate of 28 Th/s drawing 1.596 kW, while the more recent and more efficient Antminer S17 Pro has a hashrate of 53 Th/s at 2.094 kW. The corresponding mining efficiencies are 17.5 $\frac{\text{Th/s}}{\text{kW}}$ and 25.3 $\frac{\text{Th/s}}{\text{kW}}$, respectively.

Table 3.2: Blockchain Activity

	TX_N	TX_{Value}	$TX_{Value\,Coins}$	TX_{Block}	TX_{Size}	$TX_{Size\,Coins}$
Constant	0.122***	0.175*	0.055	0.105**	0.115	−0.004
	(2.84)	(1.72)	(0.62)	(2.54)	(0.92)	(−0.03)
Bitcoin	−0.009	0.061	0.075	−0.023	0.146	0.157
	(−0.09)	(0.49)	(0.68)	(−0.27)	(0.93)	(1.09)
Blackout	0.290***	0.441***	0.479***	0.222***	0.307*	0.341**
	(5.09)	(2.96)	(3.44)	(4.08)	(1.79)	(2.12)
Bitcoin×Blackout	−0.789***	−0.495***	−0.379**	0.029	−0.023	0.100
	(−6.28)	(−2.63)	(−2.10)	(0.25)	(−0.11)	(0.48)
Observations			667			

This table shows difference-in-difference regression results for blockchain activity using hourly data. TX_N is the total number of transactions recorded on the respective blockchain, TX_{Value} the USD value contained in these transactions, and $TX_{Value\,Coins}$ the same value but in the networks' native coins (Ada or Bitcoin). TX_{Block} is the average number of transactions contained in each block. TX_{Size} ($TX_{Size\,Coins}$) is the average value per transaction in USD (native coins). The variables have been standardized for each currency by subtracting their average value and dividing by their standard deviations during the week before the sample period. The treatment period of the blackout from April 16 to April 22, 2021 is compared to the 7 days before the blackout. Robust t-statistics are reported in parentheses. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

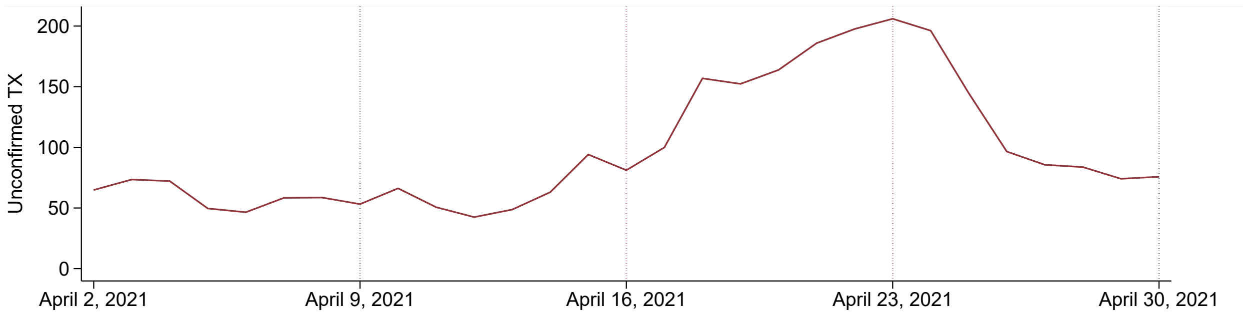
The sudden drop in hashrate confirms our first hypothesis and indicates that during the sample period covering the end of the dry season in Southwest China, a significant fraction of global Bitcoin mining is concentrated in a relatively small geographic area in northern Xinjiang and powered mostly by coal. Though Bitcoin is decentralized by design, the power outage shows limits to decentralization of PoW currencies, since miners crucially depend on low electricity prices.

3.4.2 Blockchain Activity and Transaction Fees

We continue by analyzing the activity on the currencies' blockchains in Table 3.2. The first two columns indicate that during the blackout, Ada experienced an increase in the number and USD value of the transactions recorded on its blockchain. The increases of 0.29 and 0.44 of a standard deviation relative to before the blackout are statistically significant, but substantially smaller than the corresponding decrease for Bitcoin. Compared to before the blackout, Bitcoin's transaction activity even decreases by 0.50 (0.290-0.789) and 0.05 (0.441-0.495) standard deviations for the number and USD value of transactions, respectively, confirming hypothesis 3a and consistent with a severe reduction in blockchain capacity. The average transaction size and the number of transactions included in each block increase during the blackout, confirming hypothesis 3b, showing that our results for hypothesis 3a are not completely mechanical. It is to be noted that this effect is not confined solely to Bitcoin during the blackout. We also obtain very similar results when considering the value and average transaction size denominated in the networks' native cryptocurrency tokens.

To further illustrate the congestion of the blockchain, we plot the development of the number

Figure 3.2: Number of Unconfirmed Transactions in the Mempool



This graph shows daily averages of the number of unconfirmed transactions waiting in the Bitcoin mempool, in 1,000 transactions, as provided by Hoenicke (2021). The vertical dashed lines indicate the beginning and end of the blackout and the three surrounding weeks.

of unconfirmed transactions in Figure 3.2. These transactions are waiting in the memory pool ("mempool"), i.e., they have been broadcast to the network but have not yet been included in a block by a miner. Before the blackout, the number of unconfirmed transactions is relatively stable at 60,000. During the blackout period, the number gradually increases, reaching a maximum of 218,000 at the end of the event window. The average of about 150,000 during the blackout period is statistically significantly larger than before. Afterwards, the number of transactions waiting in the mempool decreases, reverting back to its pre-event level after several days. Overall, this result is consistent with the notion that capacity on the blockchain becomes more restrictive during the blackout.

The results for the impact of the blackout and ensuing limited capacity of the Bitcoin blockchain on fees paid to miners can be found in Table 3.3. The first two columns show how the fees paid for faster settlement increase during the blackout, lending support to hypothesis 2. The descriptive statistics have shown historically high fees in the weeks before the blackout; still, the sum of all fees increases for both currencies during the event, though the effect is much stronger for Bitcoin. Since total fees incorporate both the effects of the average fees and the amount transacted, we additionally consider value-weighted relative fees. For Ada, these even slightly decrease, but for Bitcoin they soar by almost three standard deviations. The increase of 2.73 standard deviations compared to the period before the blackout translates to an economically meaningful increase in relative fees by 0.87 basis points. While still low compared to other financial assets, this constitutes an increase of about 127% relative to the average before the sample period as given in Table 3.1.

There may be substantial heterogeneity in the fees paid for individual transactions, which is not directly visible in value-weighted average relative fees. We hence analyze the distribution of equally-weighted relative fees within individual blocks. We find that the within-block dispersion of relative fees increases for both currencies. However, as with the average level of relative fees, the increase is substantially stronger for Bitcoin. Finally, while we find that Bitcoin blockchain fees increase in both the lower and upper part of their empirical distribution as given by the 10th and 90th percentile, the increase is much stronger in the upper part of the distribution.

Table 3.3: Blockchain Fees

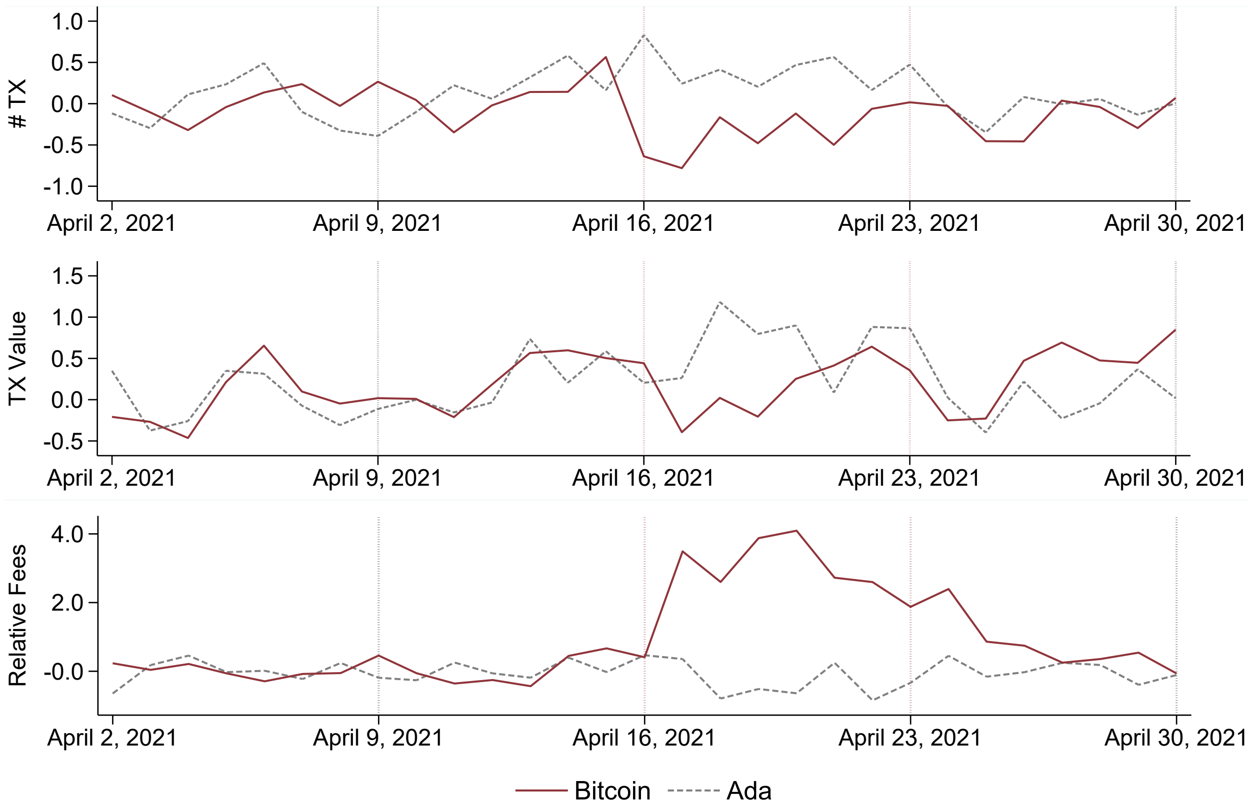
	$\text{Fees}_{\text{Total}}$	$\text{Fees}_{\text{Relative}}$	Fees_{SD}	Fees_{P10}	Fees_{P90}
Constant	0.401***	−0.010	0.589***	−0.024	0.378***
	(6.16)	(−0.17)	(6.77)	(−0.43)	(4.77)
Bitcoin	0.239*	0.075	−0.708***	0.312**	−0.216*
	(1.75)	(0.74)	(−6.70)	(2.41)	(−1.85)
Blackout	0.232***	−0.240**	0.434***	−0.016	0.165
	(2.76)	(−2.32)	(2.62)	(−0.19)	(1.28)
Bitcoin×Blackout	3.036***	2.970***	2.732***	2.143***	4.629***
	(10.73)	(11.66)	(11.45)	(9.02)	(20.29)
Observations			667		

This table shows difference-in-difference regression results for the fees paid to miners and validators using hourly data. $Fees_{Total}$ is the total amount of fees paid for the transactions in USD. $Fees_{Relative}$ is the ratio of total fees to total transaction value in basis points. $Fees_{SD}$ is the hourly average of the within-block standard deviation of equally weighted relative fees. $Fees_{P10}$ and $Fees_{P90}$ are the hourly average of the 10th and 90th percentile of within-block equally weighted relative fees, respectively. The variables have been standardized for each currency by subtracting their average value and dividing by their standard deviations during the week before the sample period. The treatment period of the blackout from April 16 to April 22, 2021 is compared to the 7 days before the blackout. Robust t-statistics are reported in parentheses. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

The results suggest that during the blackout, capacity restrictions on the Bitcoin blockchain became significantly more binding, inducing impatient traders to bid up fees. Our findings are thus consistent with the predictions of Easley et al. (2019), who found that fees reflect the traders' queuing problem, where some users choose to increase their fees while others exit in response to prolonged waiting times. Our findings of an exogenous increase in waiting times due to the blackout and a subsequent increase in fees and a reduction in the number of confirmed transactions is thus consistent with their predictions. Our results on distribution of trading fees are also consistent with strategic capacity management by miners as found by Lehar and Parlour (2022), wherein miners make processing capacity artificially scarce to extract extra fees. In addition to illustrating the impact of the blackout and subsequent reduction in blockchain capacity on trading fees and their distribution as in Lehar and Parlour (2022), our framework also allows us to compare the differential impact of this event on PoW relative to PoS networks. While we observe an overall shift in the within-block distribution of relative fees, the increase is especially severe for Bitcoin and its PoW consensus mechanism at the upper end of this distribution. Miners hence appear to be able to extract substantial rents from users with urgent trading needs, which does not appear to be the case for the cryptocurrencies in the PoS control group.

Figure 3.3 shows the development of the total number and value of confirmed transactions as well as the relative fees over time, confirming the regression results and illustrating the parallel trends before the event. In the weeks prior to the blackout, both currencies closely co-move. Starting with the blackout, the currencies diverge as Bitcoin's transactions decrease in both number and value while relative fees increase substantially. Fees do not drop immediately after the blackout, but gradually decrease back to previous levels after several days. This time frame coincides with the time it takes until the number of unconfirmed transactions waiting in the mempool decreases back to its pre-event level as seen in Fig-

Figure 3.3: Transactions and Fees over Time



The top and middle graphs show the total number and value of transactions on the blockchains, respectively. The bottom graph shows the fees relative to the value of the transactions. Both variables have been standardized by subtracting the mean and dividing by the standard deviation during the week from April 2 to April 8. The graphs show daily averages of the respective hourly variables.

ure 3.2. This suggests that miners successively pick waiting transactions from the mempool according to the offered fees, starting with those transactions offering the highest fees. While the reversion is faster for the number and especially for the value of transactions than for fees, it appears as though for Bitcoin the transaction value actually increases for some time once fees have reverted back to normal levels. This is consistent with the notion that some traders merely postpone their trading, particularly for high-value transactions.[15]

A related question is how the overall revenue of Bitcoin miners was impacted by the blackout. There are two factors working in opposite directions: Firstly, fewer block rewards are paid out because fewer blocks are generated. Since they do not participate in mining anymore, miners affected by the blackout do not receive any block rewards at all. However, the expected block rewards of unaffected miners do not change in response to the blackout since the difficulty of finding a valid block is not adjusted during the sample period. Secondly, the total fees

[15] A potential explanation for this postponement, particularly for high-value transactions, could be that traders unwilling to pay higher fees may want to avoid letting their transactions wait in the mempool for extended periods. Since all transactions waiting to be confirmed are publicly visible in the mempool, front-running poses a threat, for example when moving funds to known wallets of exchanges, indicating an intention to sell. This is a form of "miner extractable value" (see e.g. Auer et al., 2022).

paid to miners increase as shown above. Consequently, miners not directly affected by the blackout should expect higher profits while those affected incur losses.[16] Specifically, in the week before the blackout, block rewards worth USD 398mn were paid out. This decreases by USD 118mn to USD 280mn during the blackout. The increase in total fees by 3.036 standard deviations translates to USD 44mn for the entire blackout period.[17] Hence, there is an overall reduction in miners' revenues by USD 74mn.

3.4.3 Secondary Market Activity and Quality

We now turn to the effect the blackout and resulting shock to mining activity have on cryptocurrency prices and exchange trading activity. The results can be found in Table 3.4. Returns do not significantly change for either currency as seen in the first column. While several news outlets associated a concurrent decline in prices with the blackout, the observed lower returns during the blackout are within the usual volatility of the cryptocurrencies and thus insignificant. Conversely, price volatility is generally much higher during the blackout, and especially so for Bitcoin, confirming hypothesis 5. Compared to before the blackout, the volatility of Ada increases by 0.71 while Bitcoin's volatility increases by an additional 1.08 standard deviations, which in both cases is highly statistically and economically significant. Since many cryptocurrencies closely co-move with Bitcoin prices, in the third and fourth column we first calculate abnormal returns for Ada by using Bitcoin returns as a single market factor. This way, we control for any expected price changes due to Bitcoin returns. Nevertheless, our findings regarding returns do not change. However, the effect of the blackout on Ada return volatility expectedly decreases in magnitude while the effect on Bitcoin increases when using abnormal returns.

Trading volume at Kraken is higher during the blackout, especially and significantly so for Bitcoin. Consistent with the observed increase in volatility, the additional uncertainty surrounding the shock to mining and the drop in hashrate might induce traders to adjust their portfolios. Additionally and according to hypothesis 6, we find that liquidity deteriorates during the blackout as bid-ask spreads widen substantially and the Kyle and Obizhaeva (2016) illiquidity index increases. Noteworthily, liquidity decreases even though trading volume increases. The reduction in liquidity is potentially due to the increase in volatility and general uncertainty as market makers demand compensation for the additional risk due to difficulties in inventory management.

The graphs in Figure 3.4 show that volatility, volume, and estimated spreads exhibit very similar patterns. Before the blackout, both currencies again closely co-move. Compared to the changes in blockchain activity, the reaction of exchange trading activity occurs with a slightly greater delay of about one day. All measures stay at elevated levels until after about

[16]The increase in profitability of mining might incentivize some miners to switch older and less efficient hardware back on because marginal profits are now above the marginal (electricity) costs of running the hardware. However, even if this is the case, we still observe the overall drop in implied hashrate. Hence, while we cannot rule out such behavior, these miners only contribute little to the overall hashrate.

[17]This is while controlling for the overall increase in total fees paid to validators in the Cardano network.

Table 3.4: Prices and Exchange Trading Activity

	Return	Volatility	$\text{Return}_{\text{MM}}$	$\text{Volatility}_{\text{MM}}$	Volume	Spread	Illiquidity
Constant	11.516	0.461***	4.200	0.567***	0.544***	0.412***	−0.066***
	(1.21)	(4.83)	(0.49)	(5.80)	(5.29)	(4.75)	(−12.84)
Bitcoin	−6.403	−0.283**	0.914	−0.390***	−0.109	−0.299**	−0.070
	(−0.61)	(−2.18)	(0.09)	(−2.96)	(−0.71)	(−2.42)	(−0.83)
Blackout	−26.926	0.711***	−10.232	0.367**	0.191	0.370*	0.055***
	(−1.61)	(4.74)	(−0.89)	(2.38)	(1.22)	(1.94)	(6.65)
Bitcoin×Blackout	9.634	1.081***	−7.060	1.425***	0.700***	1.633***	0.775***
	(0.51)	(4.16)	(−0.48)	(5.43)	(2.61)	(4.11)	(6.63)
Observations				674			

This table shows difference-in-difference regression results for trading activity on Kraken using hourly data. *Return* is the logarithmic return of hourly closing prices in basis points. *Volatility* is the standard deviation of minutely log returns. For $Return_{MM}$ and $Volatility_{MM}$, the returns of Ada are the residuals of regressing Ada returns on Bitcoin returns. Volume is the total trading volume in USD. *Spread* is the high-low spread estimate in basis points. *Illiquidity* is the illiquidity index by Kyle and Obizhaeva (2016). The variables except returns have been standardized for each currency by subtracting their average value and dividing by their standard deviations during the week before the sample period. The treatment period of the blackout from April 16 to April 22, 2021 is compared to the 7 days before the blackout. Robust t-statistics are reported in parentheses. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

ten days, coinciding with the higher blockchain trading fees.

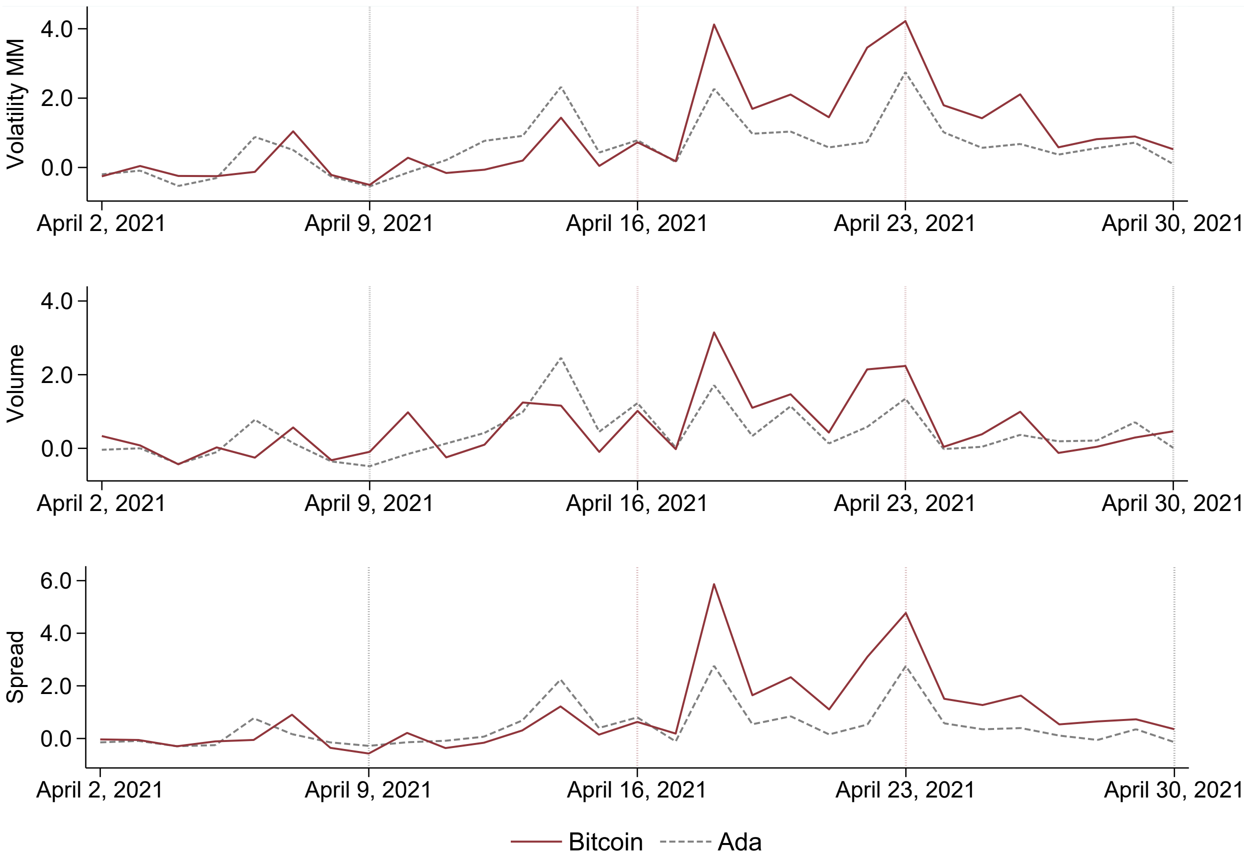
Our findings regarding exchange trading activity, especially with respect to return volatility and market liquidity, indicate that there are substantial spillover effects from mining to exchange trading activity. Operational risks faced by mining operations, like those arising from geographical mining concentration, thus have potentially severe effects on all market participants.

3.4.4 Market Integration

Finally, we study market integration by analyzing cross-venue price differences. The results in Table 3.5 and Figure 3.5 indicate that generally, price differences increase substantially. For example, price differences between Kraken and Binance increase by almost five standard deviations relative to before for both currencies. The same holds for more extreme price differences as given by their 90th percentile and for the volatility of price differences. We find very similar results for price differences between Kraken and Coinbase as well as between Binance and Coinbase, suggesting that the drop in market integration is system-wide and affects all trading venues.

Taken together, these results indicate that exploiting arbitrage opportunities during the blackout period with its higher volatility and lower liquidity becomes more difficult, confirming hypothesis 7. Many investors in cryptocurrency markets prefer holding their coins in private wallets instead of holding them directly with the exchange. (Hoang and Baur, 2022). For arbitrage trading, these investors would first have to transfer their coins to the exchange with relatively higher prices. An increase in blockchain congestion hence reduces the speed

Figure 3.4: Volatility, Volume, and Spreads



The graphs show the abnormal return volatility, trading volume, and high-low spread estimates. All variables have been standardized by subtracting the mean and dividing by the standard deviation during the week from April 2 to April 8. The graphs show daily averages of the respective hourly variables.

and profitability of exploiting mispricing. Notably, we do not observe an additional effect on Bitcoin. Rather, market integration drops for both currencies. A potential explanation for this finding is that Bitcoin is often used as a transfer currency (Kaiser and Stöckl, 2020) and hence especially important for moving funds between exchanges. Our results are also in line with Hautsch et al. (2021) and thus empirically confirm their findings regarding the negative effect of settlement latency on market integration using the exogenous shock of the blackout.

3.4.5 Robustness

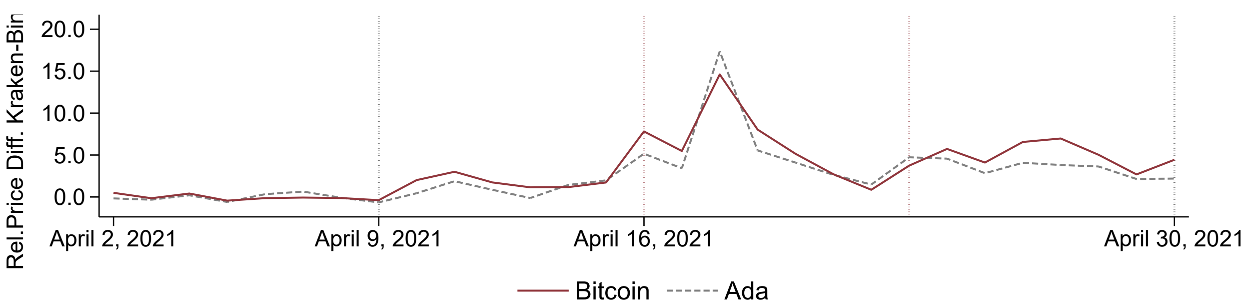
We perform a number of robustness tests. To address concerns regarding our choice of trading venue, we repeat the analysis using data from Binance, a semi-regulated exchange where Bitcoin and Ada are quoted against the stablecoin Tether, and using data from Coinbase, another large but regulated US cryptocurrency exchange. The results can be seen in Panels A and B of Table 3.6. Overall, we obtain very similar results. The major difference to Kraken

Table 3.5: Market Integration

	$\Delta P^{\text{Kraken–Binance}}_{\text{Mean}}$	$\Delta P^{\text{Kraken–Coinbase}}_{\text{Mean}}$	$\Delta P^{\text{Binance–Coinbase}}_{\text{Mean}}$	$\Delta P^{\text{Kraken–Binance}}_{\text{P90}}$	$\Delta P^{\text{Kraken–Binance}}_{\text{Std.Dev.}}$
Constant	0.834***	0.367***	0.036	0.758***	0.626***
	(7.02)	(4.74)	(0.35)	(7.04)	(6.06)
Bitcoin	0.656***	0.122	1.475***	0.356**	0.009
	(3.27)	(0.78)	(7.79)	(2.03)	(0.06)
Blackout	4.840***	4.304***	5.640***	4.330***	3.048***
	(6.14)	(2.65)	(4.12)	(6.18)	(5.41)
Bitcoin×Blackout	0.028	−2.361	−1.169	−0.806	−1.370
	(0.03)	(−1.22)	(−0.77)	(−0.89)	(−1.50)
Observations			670		

This table shows regression results for market integration as measured by the absolute value of relative price differences between Kraken, Binance, and Coinbase. The measures are based on minutely closing prices and aggregated to an hourly frequency by computing the average, the standard deviation, or the 90th percentile, respectively. The variables have been standardized by subtracting their average value and dividing by their standard deviations during the week before the sample period. The treatment period of the blackout from April 16 to April 22, 2021 is compared to the 7 days before the blackout. Robust t-statistics are reported in parentheses. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Figure 3.5: Cross-venue Price Difference



The graph shows the absolute value of relative price difference between Kraken and Binance. The variable has been standardized by subtracting the mean and dividing by the standard deviation during the week from April 2 to April 8. The graph shows daily averages of the respective hourly variable.

is that there is a stronger increase of volatility and trading volume for Bitcoin during the blackout but an overall weaker reduction in liquidity. Given the consistent results obtained across different exchanges, our findings are not contingent on any particular exchange.

While Ada is the largest non-PoW cryptocurrency during our sample period and thus can be considered as an intuitive control group by itself, we additionally verify that our results are not driven by this choice, either. Panels C and D of Table 3.6 show the same analyses as before, but while using a portfolio of non-PoW cryptocurrencies as the control currency. The equally weighted portfolio additionally contains Waves, Tronix (of the TRON network), and Atom (of the Cosmos network). Reassuringly, our results do not materially change when using the alternative control group. The only meaningful difference is that the relative drop in transaction value of Bitcoin during the blackout is not statistically significant. Overall, this yields further support to the notion that there is a decrease in market quality for Bitcoin during the blackout and that the results are not driven by idiosyncrasies of the control

Table 3.6: Robustness Tests

Panel A: *Prices and Exchange Trading Activity using Binance Data*							
	Return	Volatility	$Return_{MM}$	$Volatility_{MM}$	Volume	Spread	Illiquidity
Constant	11.463	0.474***	4.480	0.628***	0.477***	0.551***	0.175***
	(1.17)	(4.89)	(0.51)	(6.36)	(5.35)	(4.84)	(3.28)
Bitcoin	−6.470	−0.340***	0.514	−0.495***	0.023	−0.509***	−0.410***
	(−0.60)	(−2.62)	(0.05)	(−3.77)	(0.16)	(−3.69)	(−4.79)
Blackout	−26.899	0.705***	−11.073	0.166	−0.058	1.179***	1.071***
	(−1.57)	(4.66)	(−0.92)	(1.15)	(−0.55)	(7.21)	(11.92)
Bitcoin×Blackout	9.712	1.107***	−6.113	1.647***	0.867***	0.193	0.441***
	(0.50)	(3.77)	(−0.41)	(5.68)	(4.23)	(0.78)	(2.72)
Panel B: *Prices and Exchange Trading Activity using Coinbase Data*							
	Return	Volatility	$Return_{MM}$	$Volatility_{MM}$	Volume	Spread	Illiquidity
Constant	11.536	0.478***	4.495	0.612***	0.853***	0.474***	−0.240***
	(1.18)	(4.78)	(0.52)	(5.97)	(7.42)	(3.76)	(−4.03)
Bitcoin	−6.483	−0.340**	0.558	−0.473***	−0.397***	−0.306**	−0.140
	(−0.60)	(−2.53)	(0.06)	(−3.47)	(−2.68)	(−2.03)	(−1.43)
Blackout	−26.837	0.767***	−10.934	0.354*	0.263	1.419***	0.674***
	(−1.62)	(4.70)	(−0.88)	(1.93)	(1.64)	(5.96)	(7.19)
Bitcoin×Blackout	9.684	1.109***	−6.220	1.522***	0.589**	0.266	0.423***
	(0.52)	(3.59)	(−0.41)	(4.76)	(2.26)	(0.77)	(2.72)
Panel C: *Blockchain Activity using a non-PoW Portfolio*							
	TX_N	TX_{Value}	TX_{Block}	TX_{Size}	Blocktime	$Fees_{Total}$	$Fees_{Relative}$
Constant	0.066	0.055	0.018	0.054	−0.272***	0.542***	−0.024
	(1.38)	(0.89)	(0.38)	(0.91)	(−6.88)	(9.14)	(−0.69)
Bitcoin	0.048	0.181*	0.064	0.207*	0.299***	0.098	0.089
	(0.49)	(1.91)	(0.73)	(1.87)	(2.83)	(0.74)	(1.00)
Blackout	0.319***	0.161	0.296***	0.078	−0.008	−0.057	−0.044
	(4.92)	(1.55)	(4.57)	(0.78)	(−0.17)	(−0.73)	(−0.91)
Bitcoin×Blackout	−0.818***	−0.215	−0.045	0.206	0.709***	3.325***	2.774***
	(−6.32)	(−1.39)	(−0.37)	(1.25)	(3.93)	(11.83)	(11.67)
Panel D: *Prices and Exchange Trading Activity using a non-PoW Portfolio*							
	Return	Volatility	$Return_{MM}$	$Volatility_{MM}$	Volume	Spread	Illiquidity
Constant	13.932*	−0.050	−3.087	0.027	0.689***	0.013	−0.378***
	(1.78)	(−1.02)	(−0.42)	(0.57)	(6.65)	(0.33)	(−15.16)
Bitcoin	−8.819	0.227**	8.201	0.150	−0.254*	0.100	0.242***
	(−0.98)	(2.24)	(0.94)	(1.49)	(−1.65)	(1.04)	(2.77)
Blackout	−32.994*	0.640***	−8.676	0.467***	−0.008	0.241**	0.333***
	(−1.90)	(6.38)	(−0.81)	(4.76)	(−0.05)	(1.97)	(7.28)
Bitcoin×Blackout	15.702	1.152***	−8.616	1.325***	0.898***	1.763***	0.496***
	(0.80)	(4.91)	(−0.62)	(5.66)	(3.39)	(4.77)	(3.97)

This table shows robustness tests similar to Table 3.2, Table 3.3, and Table 3.4. In Panel A, exchange data from Binance is used, where Bitcoin and Ada are traded against US Tether. In Panel B, exchange data from Coinbase is used. In Panels C and D, an equally weighted portfolio of cryptocurrencies (Ada, Tronix, Atom, and Waves) using consensus protocols other than PoW is used where the exchange data comes from Kraken. *Blocktime* is the average time between two blocks in minutes. The other variables are as defined in the respective tables in the main part. Robust t-statistics are reported in parentheses. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

currency. In fact, we obtain very similar results when using each of the control currencies in the portfolio individually. It hence appears unlikely that our results are due to specific characteristics of the various control currencies.

Our results are also robust to the exact specification of the event window. While we can relatively precisely timestamp the beginning of the blackout, timestamping the end proves more difficult. For example, power may have been restored gradually to different parts of the region or miners may have taken some time to switch their operations back on. Though our estimate of the implied hashrate is consistent with a blackout duration of about one week, some of our investigated measures – like fees or volatility – stay at elevated levels longer than one week. While this does not necessarily imply a longer blackout duration, in untabulated results we empirically confirm that changing the window length by moving the end date of the blackout period by several days in either direction does not meaningfully impact our conclusions.

3.5 Conclusion

Many cryptocurrencies are designed to be decentralized, but in practice market forces may drive them towards lower degrees of decentralization. While the previous literature has focused on centralization in mining hardware or the effect of centralized mining pools, our results point to another source of centralization risk: When miners crucially depend on low electricity prices and hence accumulate in the same area, local geopolitical and operational risks can adversely affect the whole blockchain network. Since we also document strong spillover effects to exchange trading activity and secondary market quality and liquidity, these risks potentially affect a wide range of market participants. Future research investigating the welfare implications of centralization within cryptocurrencies and the relative merits of different consensus mechanisms should thus not only consider the costs and environmental externalities associated with mining, but also these indirect costs imposed on users. Likewise, traders and regulators should be aware of these systemic risks associated with energy-intensive proof-of-work cryptocurrencies.

Importantly, the shock to the network is only temporary. In our setting, this is the result of the short duration of the blackout. However, the temporary nature of mining activity shocks is also an inherent feature of the underlying proof-of-work mechanism, since mining difficulty is adjusted automatically so that changes in the overall computing power of the network only temporarily affect the speed of settlement. In the long run, this resiliency attenuates the geopolitical and operational risks that stem from the geographical centralization of mining. However, in the short term, mining shocks can still adversely impact the entire network, potentially leading to higher and more volatile fees in addition to opportunity costs due to slower settlement and missed gains from trade. While the current difficulty adjustment in Bitcoin occurs approximately every two weeks, advocating for more frequent difficulty adjustment as a policy or as a design implementation in a new upgrade can potentially mitigate the impact of mining shocks more effectively.

The power outage in one relatively small geographical region thus shows limits to decentralization of proof-of-work networks. The results additionally indicate that the Bitcoin network not only consumes vast amounts of energy, but also heavily relies on fossil fuels. Currencies based on alternative consensus mechanisms such as proof-of-stake do not necessarily share these same shortcomings stemming from the inherent dependence on electricity as the main input factor.

3.6 Appendix to Chapter 3

3.6.1 Background of Bitcoin Mining

In a decentralized payment system such as Bitcoin, the risk of double-spending, where a Bitcoin owner attempts to spend the same bitcoin twice, poses a significant concern. As a result, it is imperative that transactions in a decentralized system undergo thorough verification before being recorded to ensure their integrity. For the Bitcoin system, Nakamoto (2008b) proposes to solve the double-spending problem by verifying the transactions through the proof-of-work (PoW) consensus mechanism, a computationally demanding process.

A decentralized system comprises interconnected nodes or computers capable of sharing information. When new transactions arise, they are propagated across the network through a series of node-to-node broadcasts until reaching the majority of the network (Binance Academy, 2020b). The verifier nodes ensure the validity of the transactions, e.g. whether there is double-spending, before selecting transactions from the valid ones to form a new block. Then the verifier nodes compete to become the first to generate a 64-digit hexadecimal number adhering to a specific format, known as “hash”, from the block by using the the Bitcoin block hashing algorithm (Bitcoin Wiki, 2021). This process is referred to as “mining”, during which the hashing algorithm repeatedly hashes the block header, the part of a block that summarizes the rest of the block, along with an incrementing number called a “nonce” (Bitcoin Wiki, 2021; Coinbase, 2020; Binance Academy, 2020a).[18] Correspondingly, the transaction verifiers are referred to as “miners”.

After a valid hash is found, the miner broadcasts the newly formed block to other nodes, who would verify its authenticity.[19] If other miners accept the new block, they would try to create the next block in the blockchain, considering the accepted block as the previous block. The miners that manage to attach a valid block to the blockchain receive newly-minted bitcoins

[18]Hashrate, which measures the computing power when mining, is measured in units of hash/second, or how many hashes can be generated via Bitcoin block hashing algorithm per second (bitFlyer EUROPE S.A., 2019). New transactions may be incorporated into the block during this process (Bitcoin Wiki, 2021).

[19]A transaction is typically deemed confirmed by merchants and exchanges once six blocks have been appended to the block containing the transaction. Given an average block time of ten minutes, this implies that it takes approximately one hour to ensure that the transaction newly recorded on the blockchain is no longer at a risk of double spending. Additionally, newly-minted coins obtained by miners as part of block rewards cannot be spent for 100 blocks (Bitcoin Wiki, 2018). The block reward is halved approximately every four years (Coinbase, 2024).

(block reward) and any transaction fees paid by users, as introduced in Section 3.1. In cases of disagreement, such as when two next blocks are found simultaneously, the tie is expected to be broken after the next block is found, with the longest chain prevails and the other abandoned, i.e. becoming orphan blocks (Binance Academy, 2019). Miners always consider the longest chain as the correct one and attempt to attach new blocks to it (Nakamoto, 2008b).

As the Bitcoin blockchain grows and the amount of computational power in the network increases, miners nowadays need to combine their computing resources together to make a profit from mining, forming the mining pool (Coinbase, 2020). This trend of centralizing computing resources in the network alerts the network participants to the possibility of 51% attacks on the network, in which an entity with over 50% network computational power can theoretically undermine the network integrity by e.g. validating double-spend transactions. This fear is exemplified in 2014, when a mining pool "Gash.io" accumulated over 42% of the total Bitcoin mining power, causing concern within the Bitcoin community. However, this centralization was swiftly mitigated as miners began to exit the pool (Hajdarbegovic, 2014).

3.6.2 Summary Statistics

Table 3.7: Descriptive Statistics during the Blackout Period

	Mean	SD	P5	P50	P95	Skew.	Kurt.
Panel A: *Bitcoin*							
$\text{Transactions}_{\text{N}}$	10.35	4.56	3.35	10.07	17.36	0.4	3.0
$\text{Transactions}_{\text{Total Value}}$	4.18	2.29	1.18	3.69	8.41	0.8	3.1
$\text{Transactions}_{\text{Size}}$	0.42	0.20	0.17	0.39	0.80	1.2	5.8
$\text{Transactions}_{\text{Block}}$	2213.93	371.75	1554.25	2214.57	2872.60	0.1	3.0
Blocktime	16.42	11.66	6.46	12.76	39.38	2.5	10.3
$\text{Fees}_{\text{Total}}$	555.27	265.11	177.15	502.20	1018.31	0.6	3.0
$\text{Fees}_{\text{Relative}}$	1.59	0.90	0.64	1.32	3.23	1.6	6.8
$\text{Fees}_{\text{SD within Block}}$	78.03	28.51	38.07	74.19	131.18	1.2	5.1
$\text{Fees}_{\text{P10 within Block}}$	0.04	0.02	0.01	0.03	0.07	0.6	2.5
$\text{Fees}_{\text{P90 within Block}}$	24.26	5.22	15.16	24.58	31.50	0.1	4.0
Return	−11.96	98.90	−186.14	−2.78	134.20	−1.1	8.6
Volatility	11.53	6.72	4.83	9.84	22.40	4.1	31.0
Volume	17.02	16.17	5.25	13.31	39.55	4.9	37.1
Spread	1.40	1.86	0.40	0.97	2.99	6.5	53.3
Illiquidity	1.59	0.34	1.09	1.57	2.25	0.4	3.1
Cross-venue Price $\text{Diff.}_{\text{Mean}}$	11.87	8.66	4.08	10.53	20.95	4.9	35.8
Cross-venue Price $\text{Diff.}_{\text{P90}}$	17.34	15.25	8.42	14.96	26.22	7.4	64.2
Cross-venue Price $\text{Diff.}_{\text{Std.Dev.}}$	4.37	6.66	2.82	3.54	5.70	10.2	112.5
Panel B: *Ada*							
$\text{Transactions}_{\text{N}}$	1.79	0.35	1.28	1.79	2.47	0.6	3.5
$\text{Transactions}_{\text{Total Value}}$	0.31	0.25	0.11	0.30	0.42	4.9	31.1
$\text{Transactions}_{\text{Size}}$	0.18	0.16	0.05	0.16	0.28	4.8	31.7
$\text{Transactions}_{\text{Block}}$	11.02	2.13	7.75	11.08	14.84	0.5	3.4
Blocktime	0.37	0.03	0.33	0.37	0.42	0.3	3.7
$\text{Fees}_{\text{Total}}$	0.50	0.11	0.34	0.49	0.76	0.7	3.3
$\text{Fees}_{\text{Relative}}$	0.02	0.01	0.01	0.02	0.05	1.3	4.4
$\text{Fees}_{\text{SD within Block}}$	1.12	0.49	0.61	1.00	2.33	1.7	5.5
$\text{Fees}_{\text{P10 within Block}}$	0.05	0.07	0.00	0.02	0.16	4.5	28.0
$\text{Fees}_{\text{P90 within Block}}$	1.46	0.62	0.76	1.33	2.96	1.3	4.6
Return	−15.24	178.47	−304.00	−7.51	253.70	−0.8	7.0
Volatility	20.84	12.12	11.21	18.40	38.13	5.6	49.9
Volume	2.23	2.22	0.60	1.58	5.91	4.6	35.5
Spread	1.63	2.39	0.29	1.10	3.79	7.0	63.4
Illiquidity	6.24	1.68	3.73	6.02	9.09	0.9	4.5
Cross-venue Price $\text{Diff.}_{\text{Mean}}$	16.32	17.87	8.21	13.20	25.60	7.3	64.0
Cross-venue Price $\text{Diff.}_{\text{P90}}$	28.54	28.30	15.79	23.52	38.54	6.6	49.7
Cross-venue Price $\text{Diff.}_{\text{Std.Dev.}}$	9.23	9.00	5.70	7.55	12.54	8.1	76.4

This table shows summary statistics based on hourly data from the treatment period of April 16 to April 22, 2021. $Transactions_N$ is the number of transactions recorded on the blockchain in 1k, $Transactions_{Total\ Value}$ their value in USD 1bn, and $Transactions_{Size}$ their average size in USD 1mn. $Transactions_{Block}$ gives the average number of transactions per block and *Blocktime* the average time between two blocks in minutes. $Fees_{Total}$ is the sum of all fees paid by users in USD 1k and $Fees_{Relative}$ the same relative to the value of the transactions in basis points. $Fees_{P10/P90\ within\ Block}$ is the hourly average of the 10th/90th percentile of equally-weighted relative fees within a block in percentage points and $Fees_{SD\ within\ Block}$ the hourly average of the standard deviation of equally-weighted relative fees within a block in percentage points. The remaining variables are based on trading data from Kraken: *Return* is the logarithmic return of hourly closing prices in basis points, *Volatility* the standard deviation of minutely log returns in basis points, *Volume* is the trading volume in USD 1mn, *Spread* is the high-low spread estimate in basis points, and *Illiquidity* is the illiquidity index by Kyle and Obizhaeva (2016) in basis points. *Cross-venue Price* $Diff._{Mean/P90/Std.Dev.}$ is the hourly average/90th percentile/standard deviation of relative USD price difference between the exchanges of Kraken and Binance in basis points.

Chapter 4

Intraday Herding and Attention around the Clock[1]

4.1 Introduction

Herding in financial markets describes the inclination of individual investors to mimic the investment decisions of other investors instead of trading on their private information. Various factors contribute to such behavior, encompassing both rational motives and potential indicators of behavioral biases. Documented in practically all financial markets, herd behavior might lead to inefficient prices as investors disregard fundamental information, thereby creating irrational bubbles. Understanding such behavior and its underlying drivers is thus important for investors and regulators alike.

In this paper, we analyze investor herding at the intraday level in the decentralized cryptocurrency market, where herding behavior is particularly interesting. Contrary to most other markets, cryptocurrency markets are decentralized and open around the clock, allowing for an analysis of herding patterns throughout the day. Additionally, because there is relatively little fundamental information available or, where available, might be complex to evaluate due to the novelty of the assets, there are potentially higher levels of herding as investors follow the market instead of relying on coin-specific information. Similarly, because the market for cryptocurrencies is still young and developing, price inefficiencies might be more pervasive than in other markets. Finally, with a large fraction of retail traders

[1] This chapter has a published version in the Journal of Behavioral and Experimental Finance Scharnowski and Shi (2024). This work is the result of collaboration with Dr. Stefan Scharnowski. As a group of co-authors, we thank Peter Albrecht (discussant), J. Anthony Cookson, Hossein Jahanshahloo (discussant), Jérémy Morvan (discussant), Susanne Siedhoff (discussant), Erik Theissen, and conference and seminar participants at the Annual Meeting of the French Finance Association 2022, the Behavioural Finance Working Group Annual Conference 2022, the CryptoAssets and Digital Asset Investment Conference Rennes 2022, the Crypto Asset Lab Conference 2021, the Cryptocurrency Research Conference 2021, the World Finance Conference 2021, and the University of Mannheim for helpful comments and suggestions. We gratefully acknowledge support by the Julius-Paul-Stiegler-Gedächtnisstiftung.

(Dyhrberg et al., 2018), the investor base of cryptocurrencies is different from other, more mature asset classes, making an analysis of how investor attention affects herding especially relevant.

While there is already a rich literature on herding in cryptocurrencies, there is no consensus on the size or even presence of such behavior. We make several contributions: Firstly, using intraday data allows us to detect short-term herding by investors, which might go undetected when using lower frequencies as pointed out by Gleason et al. (2004). Moreover, the higher frequency of our dataset allows us to look at variations in herding behavior throughout the day. Secondly, we provide evidence on how investor attention and its cross-sectional dispersion affect herding while controlling for general trends in the market. To the best of our knowledge, we are the first to apply the concept of attention dispersion to herding in any financial market. Thirdly, we relate herding to both on- and off-blockchain trading activity.

In our empirical analysis, we first document substantial herding behavior that is stronger when market returns are positive, which we attribute to a fear-of-missing-out during times of increasing prices and short-sell restriction during times of decreasing prices. Herding is negatively related to both the level and cross-sectional dispersion of investor attention, suggesting that these are also important factors for herd behavior. Moreover, we uncover pronounced intraday variations in investor herding. Investor herding exhibits analogous intraday patterns to the intraday variations of attention and blockchain activity and is strongest during the overlap of hours when traders in major economic centers are likely awake.

This paper contributes to multiple streams within the literature: We add to the literature on investor herding, in particular that of retail investors, by providing novel evidence on the intraday patterns of herding in a global and continuously open market. By focusing on the cryptocurrency market, we also contribute to the understanding of price formation and investor behavior in this growing market. Additionally, we relate to the literature on investor attention by linking intraday herding to both the level and dispersion of investor attention.

Herding behavior and informational cascades have been studied extensively both theoretically and empirically across many different markets and asset classes. In their seminal empirical works, Christie and Huang (1995) and Chang et al. (2000) suggest using measures of the dispersion of returns in relation to the market return to capture herd behavior. The general idea of their measures is that less dispersion of individual asset returns around the market return indicates investor herding across assets. While neither study uncovers significant evidence of herding in the United States, Chang et al. (2000) document substantial evidence of herding in the two emerging markets within the sample: South Korea and Taiwan. Analyzing equity markets in 35 countries, Chen et al. (2022) show that around the time of earnings disclosures, stronger herding is associated with greater price informativeness. Regarding other asset classes, investor herding has been analyzed for exchange traded funds (Gleason et al., 2004), foreign exchange (Park, 2011; Sibande et al., 2023), corporate bonds (Cai et al., 2019), options (Bernales et al., 2020), and commodities (Youssef, 2022a; Liu et al., 2023). Furthermore, there is evidence that herding is a global phenomenon (Chiang and Zheng, 2010) and differs between sectors and industries (Choi and Sias, 2009; Gebka and

Wohar, 2013). Herding behavior is not limited to retail investors (Hsieh et al., 2020; Zheng et al., 2021), but has also been documented for institutional investors (Sias, 2004; Kremer and Nautz, 2013). While most previous studies measure herding at a daily frequency, some analyze herding using intraday data (Gleason et al., 2004; Hsieh, 2013; Andrikopoulos et al., 2017; Cai et al., 2019; Wang et al., 2022).

The literature looking specifically at herding in cryptocurrencies so far does not agree on the prevalence of the behavior. Bouri et al. (2019), Youssef (2022b), and Yousaf and Yarovaya (2022) find significant but time-varying herding behavior. da Gama Silva et al. (2019) find overall weak evidence of herding. Depending on the methodology, herding is found to be more significant during down markets. Vidal-Tomás et al. (2019) observe that, while extreme price movements can generally be explained by a rational asset pricing model, there is significant herding during down markets. By considering sentiment expressed in online forums, Gurdgiev and O'Loughlin (2020) find cryptocurrency-market specific herding which is more prevalent in bear markets. In contrast, Papadamou et al. (2021), Kallinterakis and Wang (2019), and Ballis and Drakos (2020) conclude that herding is more pronounced during up markets. Using a large cross-section of cryptocurrencies, Kaiser and Stöckl (2020) discover strong evidence for herding in both bull and bear markets. Closely related to our study, Philippas et al. (2020) look at how potentially informative signals affect herding activity. They find heterogeneity in how information from various external factors is taken into account by investors. For example, media attention related to Bitcoin increases herding, while high equity returns are associated with reduced herding in the cryptocurrency market. As in other markets, most studies on herding in cryptocurrency markets focus on daily data. While there are some studies that use intraday data (Yarovaya et al., 2021; Mandaci and Cagli, 2022; Choi et al., 2022; Mohamad and Stavroyiannis, 2022; Blasco et al., 2022), none focus on the intraday patterns of herding behavior or how these relate to investor attention.

Our paper also relates to the literature on intraday patterns in cryptocurrency trading activity and on the importance of using higher-frequency data. For example, Dyhrberg et al. (2018) and Eross et al. (2019) document intraday patterns in cryptocurrency trading activity that resemble those found in foreign exchange markets. They additionally find significant intraday patterns in both volatility and liquidity. Hu et al. (2019) show that price clustering at round numbers is relatively stable throughout the day. Baur et al. (2019) do not find substantial intraday patterns in returns but in the trading volume of various exchanges. Petukhina et al. (2021) find intraday patterns in volatility and trading volume that are not consistent with a full automation of trading by algorithms. Instead, they conclude that much of trading is driven by human traders. Brauneis et al. (2023) document intraday patterns in trading activity, liquidity, and volatility. These patterns are similar to each other and across exchanges located in different geographic regions, suggesting that they are partially explained by common global factors. Finally, Aslan and Sensoy (2020) highlight that conclusions regarding the efficiency of cryptocurrency prices depend on the sampling frequency, which further motivates our study.

Investor herding may be related to limited attention resulting from cognitive limitations in information processing so that investors only exhibit "approximate rationality" (Simon, 1955, 1956). Herding around the market consensus might then be a heuristic to simplify the investment decision process as in Tversky and Kahneman (1974). Investor herding is particularly related to “illusory correlation”, i.e., the overestimation of the frequency of the co-occurrence of events closely related to each other, and the resulting overconfidence in prediction. Heuristic learning in financial markets is also discussed by Hirshleifer (2015), in which the tendency of extrapolation is particularly related to investor herding. Moreover, Hirshleifer and Teoh (2003) examine the consequences of limited attention regarding firm disclosures and its impact on market prices.

We link investor herding to investor attention. Since attention is generally not directly observable, many proxies for investor attention have been proposed in the literature, for example based on extreme returns (Barber and Odean, 2008). However, a more direct proxy for investor attention and deliberate information demand is given by internet search volume. In early work, Da et al. (2011) use Google search volume to proxy for investor attention and find evidence that it likely captures the attention of less sophisticated retail traders and helps predict price movements in the following weeks. Building on these results, Joseph et al. (2011) also use Google search volume to find that it reflects buy pressure by retail traders. Similar to our study, Meshcheryakov and Winters (2020) do not rely on Google search volume at daily or lower frequencies but instead use hourly data. Higher search activity is followed by increased trading volume and smaller order sizes. They posit that the increase in trading activity is driven by retail traders who mistakenly think they are informed.

Whereas internet search volume captures the informational demand aspect of investor attention, the supply of new information might also affect herding behavior. A particularly active platform of information exchange in the context of cryptocurrencies are internet message boards (Phillips and Gorse, 2018). Antweiler and Frank (2004) analyze messages posted to two internet stock message boards and find that messages regarding particular stocks lead to increases in volatility. The more strongly the content of different messages disagrees, the larger the subsequent increases in trading volume. Sabherwal et al. (2011) analyze pump-and-dump behavior related to message board posts when there is no new fundamental information. Their results suggest that message boards can be used to induce investor herding to drive up prices.

With a substantial fraction of retail traders and some extreme price movements in the past, cryptocurrency returns tend to be particularly affected by investor attention. Phillips and Gorse (2018) proxy for investor attention by considering various online and social media attention proxies. Among them are Google search volume and posts and comments on Reddit, a message board popular among cryptocurrency traders. They find particularly strong correlations between the factors and prices during periods of bubble-like price increases. Using internet search volume as an attention proxy, Piccoli and Chaudhury (2019) demonstrate that the significant speculative bubbles observed in Bitcoin and Ethereum are driven by pronounced price surges, which capture the attention of noise traders—largely uninformed retail investors. This phenomenon contributed to herding behavior during the bubble. Zhang and Wang (2020) find that high investor attention is associated with positive returns. Similarly, Jafarinejad and Sakaki (2018) show that Bitcoin-related search volume is significantly positively related to the conditional volatility of Bitcoin returns. Philippas et al. (2020) document that Bitcoin-related tweets and Google search volume amplify investor herding. Similarly, Gurdgiev and O'Loughlin (2020) use sentiment expressed on Bitcointalk.org forum to find presence of cryptocurrency-market specific herding.

While most studies focus on the level of investor attention, only a limited number of them examine the cross-sectional relationships of attention across individual assets. Drake et al. (2017) introduce the concept of attention co-movement, which measures the extent to which firm-specific attention is related to the attention paid to the industry or to the whole market. They then show that the co-movements of attention and of returns are positively related. Similarly, See-To and Yang (2017) consider investor sentiment dispersion, which is measured using textual analysis of tweets that contain stock tickers. While sentiment dispersion does not appear to affect future returns, there is an almost immediate increase in realized volatility which then decreases during the subsequent days. To the best of our knowledge, our study is the first to relate the concept of attention dispersion to the context of investor herding.[2]

We conclude that the literature linking attention to herding is scarce, particularly when it comes to attention dispersion. Furthermore, most previous studies investigate investor attention at lower frequencies. While these studies do establish the influence of investor attention on various aspects of asset prices, there is a possibility that some of the finer dynamics in the intricate interplay between attention and trading behavior have been overlooked. We attempt to fill that gap by investigating the investor attention at hourly frequency to uncover the intraday patterns of the investor attention and how it influences investor herding.

The remainder of this paper proceeds as follows: Section 4.2 develops our hypotheses. Section 4.3 describes the dataset and the empirical approach. Section 4.4 discusses the results for herding and its intraday patterns before presenting some robustness tests, while Section 4.5 concludes.

4.2 Hypotheses

In this part we develop the hypotheses which are then tested below. There are several reasons to expect that cryptocurrency investors exhibit herding behavior. With relatively little fundamental information available, it is likely that investors follow the market more strongly than in conventional financial markets. Similarly, because the market for cryptocurrencies is still developing, there might be stronger price inefficiencies than in other markets. Finally, with a large fraction of retail traders, the investor base of cryptocurrencies is different from other, more mature asset classes (Dyhrberg et al., 2018). We therefore hypothesize that investor herding is prevalent in the cryptocurrency market.

Hypothesis 1: Cryptocurrency investors show herding behavior.

[2] We note that the literature on cross-sectional differences in attention is also closely related to the more general question of disagreement between market participants, see e.g. Carlin et al. (2014).

Previous studies have shown that the strength of herding behavior is not constant, but rather conditional on the market environment (see e.g. Chang et al., 2000; Kallinterakis and Wang, 2019; Raimundo Júnior et al., 2022; Wang et al., 2022; Vidal-Tomás et al., 2019). While it is mostly hypothesized that herding is more pronounced during times of market stress, the same might not hold for cryptocurrency markets. In particular, a fear-of-missing-out might induce traders to herd during times of extreme price increases (see e.g. Piccoli and Chaudhury, 2019). We are hence agnostic on the direction of the effect and consider this to ultimately be an empirical question.

Hypothesis 2a: Herding is asymmetric and stronger when prices are increasing.

Hypothesis 2b: Herding is asymmetric and stronger when prices are decreasing.

The demand for both on- and off-blockchain transactions in cryptocurrencies likely fluctuates over time.[3] In the long run, it should be correlated with the size of the investor base and the popularity of the currency, whereas any short-term fluctuations might reflect trading based on newly available information or stem from arbitrage activities. However, high levels of transaction activity might also result from speculation and the formation of bubbles. We hence hypothesize that both exchange trading volume and the number of transactions recorded on the blockchain are related to herding activity, but the direction of the effect is, again, an empirical question.

Hypothesis 3a: Trading volume and blockchain transaction activity are positively to herding behavior.

Hypothesis 3b: Trading volume and blockchain transaction activity are negatively to herding behavior.

High levels of investor attention might be associated with lower market herding because investors seek — and find — more private information. This especially holds when attention is measured via internet search volume since this is a direct proxy for informational demand. Likewise, informational supply as measured by posts on internet message boards is positively associated with investor attention. While we expect that higher levels of aggregate attention already have a negative effect on herding, we anticipate an additional negative effect when the cross-sectional dispersion of attention is high. The reason is that high dispersion indicates that attention is directed towards specific currencies and does not solely reflect an increase in interest in cryptocurrencies in general.

Hypothesis 4: Herding is negatively related to both the level and the dispersion of investor attention.

[3] Depending on the cryptocurrency, on-blockchain transactions include for example payments for goods and services, but more commonly settlement of transactions at centralized exchanges or transactions related to decentralized applications such as decentralized exchanges, lending and borrowing, or gaming.

Contrary to most other financial markets, cryptocurrencies trade around the clock, allowing for an analysis of differences in herding behavior throughout the day. We hence hypothesize that herding behavior varies throughout the day, but interpreting the exact pattern is complicated by the fact that the cryptocurrency market is global, decentralized, and anonymous, so that it is unclear in which time zones traders are located. However, because prior studies have documented distinct patterns in global cryptocurrency trading activity and liquidity, we expect herding to be most prevalent when global activity is strongest. According to Brauneis et al. (2023), this would be the afternoon in Coordinated Universal Time (UTC).

Hypothesis 5: Herding activity varies throughout the day and is strongest when global trading activity is high.

4.3 Methodology and Data

4.3.1 Cryptocurrency Data

We obtain hourly intraday data on 12 cryptocurrencies: Bitcoin (BTC), Cardano (ADA), Dash (DASH), Dogecoin (DOGE), Ethereum (ETH), Ethereum Classic (ETC), Litecoin (LTC), Monero (XMR), Ripple (XRP), Stellar (XLM), Tronix (TRX), and Zcash (ZEC). All prices are in USD. The sample spans from July 1st, 2017, to March 31st, 2022. At the beginning (end) of our sample, the included cryptocurrencies represent more than 85% (75%) of the total cryptocurrency market capitalization.[4]

Data quality and reliability is a particular concern when analyzing cryptocurrency markets. For example, Alexander and Dakos (2020) warn that using non-traded prices from so-called "coin-ranking" websites might lead to inconsistent results. Our sample is hence based on trade data from Kraken, which has been identified as one of the trustworthy crypto exchanges (Härdle et al., 2020). For example, there is no evidence that it reports inflated trading volume.[5]

We then calculate logarithmic returns for cryptocurrency i at time t based on hourly closing prices and, similarly to Chang et al. (2000), use these to construct an equally weighted

[4]Our choice of sample cryptocurrencies is motivated by their listing at the exchange Kraken and the length of available data. Furthermore, we exclude stablecoins.

[5]However, to additionally verify that the data is representative of the overall cryptocurrency market and that our results do not depend on our specific data source, we compare it to prices determined by coinmarketcap.com. Differences are generally small: The average (median) difference between these two prices is 0.52% (0.37%) and similar across the various currencies, though some price differences between exchanges are expected due to differences in trading fees and liquidity. To filter any remaining outliers, we drop observations where the absolute difference is larger than 10%.

market portfolio:

$$R_{i,t} = \ln\left(\frac{C_{i,t}}{C_{i,t-1}}\right) \qquad R_{m,t} = \frac{1}{N_t}\sum_{i=1}^{N_t} R_{i,t}$$

For robustness we also use a value weighted index and a market index solely based on Bitcoin returns.[6] As in Kaiser and Stöckl (2020), we allow the number of included cryptocurrencies N to change over time, in particular as newer cryptocurrencies enter the sample. We require the market index to be based on at least five currencies at each point in time, but typically the number is substantially larger. On average, about ten of the twelve cryptocurrencies are part of the market index, and more than 95% of the time there are at least eight.

We additionally obtain the number of transactions recorded on the blockchain of each cryptocurrency in the sample. The data is collected by connecting to publicly available APIs for the various currencies. The number of transactions contained in every block is counted and aggregated to one-hour intervals to match the trading data. We then normalize the transaction data by winsorizing at the 99.5% level, dividing each time series by their respective maximum transaction count, and multiplying the result by 100. Analogously to the attention measure below, we then aggregate the individual transaction counts to the market level:

$$\text{BlockchainTransactions}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} \ln\left(1 + \text{Tx}_{i,t}\right)$$

4.3.2 Investor Attention and Information Demand and Supply

Following Da et al. (2011), Joseph et al. (2011) and Fink and Johann (2014), we use the search volume index (SVI) from Google Trends to measure investor attention and information demand. As search keywords, we use the full name of the cryptocurrency, unless the name does not unambiguously return results related to the cryptocurrency. In those cases, we use the ticker or add the word “coin” to the name of the cryptocurrency. The keywords are thus Bitcoin, Cardano, Ethereum, Ethereum Classic, Dash coin, Dogecoin, Litecoin, Monero, Ripple, Tron coin, XLM, and Zcash.[7] Our measure of market-wide search activity and thus attention is given by the average of the logarithms of relative search volume across

[6] We can not utilize CRIX for this purpose due to its historical data being available only at a daily frequency.

[7] Google Trends only returns hourly data for relatively short time spans. For a given keyword-timeframe combination, the raw data is always expressed relative to the highest search volume in that timeframe which is set to a value of 100. The other relative values are rounded to the nearest integer and set to zero if below an unknown threshold. To obtain a long hourly sample with consistent scaling in the time series, we start with the first week of the sample and then move forward in time by six days, leaving 24 observations per keyword as an overlap which we use to consistently scale the data in the time series. Finally, we winsorize the data at the 99.5% level, divide every time series by the maximum SVI of the respective currency, and multiply by 100.

all currencies that are part of the market portfolio for a given hour:

$$\text{SearchVolumeLevel}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \ln\left(1 + S_{i,t}\right)$$

Additionally, we measure the cross-sectional dispersion of investor attention similarly to the return dispersion measure below by taking the average absolute deviation from the market search volume level:

$$\text{SearchVolumeDispersion}_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \left|\ln\left(1 + S_{i,t}\right) - \text{SearchVolumeLevel}_t\right|$$

We further obtain data on messages posted on Reddit as a proxy for information supply. The platform is especially popular with retail and cryptocurrency investors (Phillips and Gorse, 2018). We collect all submissions and comments (called posts henceforth) from the Pushshift archives (Baumgartner et al., 2020). We then count the number of posts on the respective main message board (“Subreddit”) for each cryptocurrency. The data is aggregated to market-wide measures of the level and cross-sectional dispersion of the number of Reddit posts analogously to the internet search volume measures.[8]

4.3.3 Measuring Herding Behavior

Following Chang et al. (2000) and many subsequent studies on investor herding, we consider the cross-sectional absolute deviation (CSAD) of individual cryptocurrency returns from the market return, i.e. the mean of market model residuals of cryptocurrencies in the sample.

$$\text{CSAD}_t = \frac{1}{N} \sum_{i=1}^{N} \left|R_{i,t} - R_{m,t}\right|$$

While CSAD per se is not a measure of herding around the market consensus, but a non-linear relationship between this measure and the market return may indicate the presence of herding behavior. To formally test this notion, Chang et al. (2000) suggest regressing the CSAD on absolute and squared market returns. If investors exhibit return herding behavior, we expect a significantly negative coefficient for the squared market returns. The intuition is that a rational and linear asset pricing model such as the CAPM would predict a linear relationship between return dispersion and market returns. However, if there is herding

[8] A noteworthy point about the measures of attention and information supply is that they are based on data that is normalized in the time series. Dispersion hence does not measure difference in absolute search volume or Reddit posts, but rather differences in relative values. We employ this approach to address the vastly different levels of attention the larger cryptocurrencies such as Bitcoin receive compared to some of the smaller altcoins. This approach is thus consistent with using an equally weighted market portfolio.

around the market consensus during periods of market stress, return dispersion will decrease in the market return or at least increase at a decreasing rate.[9]

Formally, we use various specifications of the following regression equation:

$$\begin{aligned}\text{CSAD}_t = \alpha &+ \beta_1 \left|R_{m,t}\right| + \beta_2 R_{m,t}^2 \\ &+ \beta_3 \text{TradingVol}_t + \beta_4 \text{BlockchainTrans}_t \\ &+ \beta_5 \text{SearchVol}_t + \beta_6 \text{RedditPosts}_t \\ &+ \sum\nolimits_d \tau_d D_{(t\in d)} + \varepsilon_t\end{aligned} \tag{4.1}$$

where $D_{(t\in d)}$ indicates the date d and is used to capture date-fixed effects. We use this equation to test for the overall presence of herding behavior in intraday returns, which are likely more sensitive to short-lived herding.

Firstly and as a baseline specification, we only include the market return measures. While the coefficient for absolute market returns is of lesser interest and primarily controls for the expected return dispersion under a rational asset pricing model such as the CAPM, the coefficient for squared market returns indicates herding behavior when significantly negative. Importantly and as mentioned above, this approach captures herding around the market consensus, i.e., around the market return, during times of market stress, i.e., when returns are large in magnitude.

We then include additional variables in the model. This approach follows Bernales et al. (2020) who postulate that additional variables (called "herd variables" in their study) included in such a regression should not have an impact on return dispersion as measured by CSAD under the null hypothesis of no herding. The authors consider significant coefficients for these additional variables a case of "conditional herding". Specifically, we then include variables related to trading volume and blockchain activity to see if herding is impacted by the level of both off- and on-blockchain activity. *BlockchainTrans* captures the transaction activity in the currencies of the market portfolio. *TradingVol* is the hourly trading volume at Kraken across all included cryptocurrencies. While this variable only captures a fraction of global trading activity in cryptocurrencies, it still proxies for the trading intensity of investors using US Dollars, especially when we consider intraday variations. Furthermore, Brauneis et al. (2023) show that trading activity is generally highly correlated across exchanges and geographical regions. In the next step, we include variables related to investor attention: *SearchVol* is a vector containing the measures for the level and dispersion of information

[9] Note that Bohl et al. (2017) and Stavroyiannis et al. (2019) present evidence that the methodology of Chang et al. (2000) likely underestimates the presence and magnitude of herding behavior. The reason is that when using realized returns with idiosyncratic components as opposed to expected CAPM returns, under the null hypothesis of no herding, the true coefficient for the squared market returns β_2 is actually expected to be some positive — but generally unknown — value. By testing against a coefficient of zero, one underestimates herding and overestimates anti-herding behavior, which has frequently been documented in the empirical literature on herding (e.g. Bouri et al., 2019; Coskun et al., 2020). We never document any anti-herding behavior, so in our study this bias can only work against finding significant return herding. Since in almost all of our specifications we in fact find significant herding even when testing the coefficient against zero, we conclude that while we might underestimate the magnitude of herding, our general conclusions are not affected by this bias.

demand as measured by internet search volume and *RedditPosts* similarly contains measures for the level and dispersion of information supply as measured by submissions and comments on the message board Reddit. Finally, the use of intraday data allows us to add fixed effects for every day in the sample, which controls for general trends in the data and thus focuses the analysis on shorter-term variations in herding behavior.

Following Chang et al. (2000), we repeat these regressions separately for up and down markets to study any asymmetry in herding behavior. As is commonly done in the literature, we use Newey and West (1987) standard errors to account for the potential autocorrelation in the residuals. We set the number of lags equal to $\lceil T^{0.25} \rceil$ where T is the number of observations in the regression. In the regressions with date fixed effects, we cluster the standard errors by date, though the exact choice of standard errors does not seem to meaningfully impact our results.

Previous studies have shown that return herding behavior may be time-varying (Bouri et al., 2019; Yarovaya et al., 2021). The intraday data allows us to test for another type of time-variation: Patterns in intraday herding behavior. To investigate such patterns, we estimate

$$\mathrm{CSAD}_t = \alpha + \sum_{h=0}^{23} \beta_{1,h} \left| R_{m,t} \right| D_{h,t} + \sum_{h=0}^{23} \beta_{2,h} R_{m,t}^2 D_{h,t} + \varepsilon_t \tag{4.2}$$

where D_h is a set of binary variables for each one-hour window of the day. The vector of regression coefficients β_2 then shows how herding behavior fluctuates throughout the day.

4.4 Results

4.4.1 Summary Statistics

In Table 4.1 we provide summary statistics, first for the total sample and then split into periods where the market return is positive or negative, denoted as respectively up (Panel B) and down market (Panel C). The average cross-sectional absolute deviation of hourly returns around the market is about 48 basis points. For comparison, the (untabulated) time series average absolute deviation of the hourly market return from its mean is about 60 basis points. Market return volatility and cross-sectional dispersion are thus both economically meaningful and at comparable levels. Both show signs of fat tails with substantial excess kurtosis.

While on average return dispersion is similar in up and down markets, there are more extreme values when prices are increasing as evidenced by the larger maximum and minimum values, respectively, and the higher excess kurtosis. We observe a similar pattern for trading volume: While still similar on average during up and down markets, it exhibits more extreme values for up markets. Market returns are more volatile during down markets, suggesting the presence of asymmetric volatility. The measures for attention and blockchain activity behave quite similarly during both market states.

Untabulated augmented Dickey-Fuller tests reject the null hypothesis of a unit root for all variables. Likewise, multicollinearity does not appear to pose a problem as all variance inflation factors are well below five in any of the estimated models below.

In Table 4.2 we provide correlations between the variables included in our analysis, as defined in Table 4.1. The highest correlation is the one between the level of information demand as measured by google search volume and the level of information supply as measured by the number of submissions and comments on Reddit, which is 0.73. This is another evidence that the multicollineraity does not appear to pose a problem in our analysis.

Table 4.1: Descriptive Statistics

	Mean	**SD**	**Min**	**P5**	**P50**	**P95**	**Max**	**Skew.**	**Kurt.**	**N**
Panel A: *Full Sample*										
CSAD	0.48	0.31	0.00	0.18	0.39	1.06	7.75	2.7	20.5	40,799
Market Return	−0.00	0.94	−10.81	−1.45	0.02	1.33	11.29	−0.6	13.2	40,799
Trading Volume	7.92	13.60	0.00	0.56	3.22	30.99	361.44	6.3	85.4	40,799
Blockchain Transactions	3.02	0.42	1.70	2.39	2.95	3.72	4.22	0.3	2.1	40,799
Search Volume$_{\text{Level}}$	2.00	0.69	0.43	1.09	1.88	3.35	4.55	0.8	3.2	40,799
Search Volume$_{\text{Dispersion}}$	0.82	0.22	0.07	0.42	0.84	1.17	2.13	−0.2	3.2	40,799
Reddit Posts$_{\text{Level}}$	2.13	0.59	0.40	1.23	2.12	3.15	4.62	0.3	2.8	40,799
Reddit Posts$_{\text{Dispersion}}$	0.66	0.16	0.04	0.40	0.65	0.92	1.55	0.2	3.4	40,799
Panel B: *Up Markets*										
CSAD	0.49	0.32	0.01	0.18	0.40	1.10	7.75	2.8	24.0	21,107
Market Return	0.58	0.64	0.00	0.03	0.38	1.77	11.29	3.2	23.1	21,107
Trading Volume	7.87	13.18	0.00	0.58	3.29	30.64	361.44	6.6	101.4	21,107
Blockchain Transactions	3.03	0.42	1.70	2.39	2.97	3.72	4.22	0.2	2.1	21,107
Search Volume$_{\text{Level}}$	2.01	0.70	0.43	1.09	1.89	3.36	4.55	0.7	3.2	21,107
Search Volume$_{\text{Dispersion}}$	0.82	0.22	0.07	0.42	0.84	1.17	1.74	−0.2	3.2	21,107
Reddit Posts$_{\text{Level}}$	2.13	0.59	0.64	1.24	2.12	3.15	4.62	0.3	2.8	21,107
Reddit Posts$_{\text{Dispersion}}$	0.66	0.16	0.04	0.40	0.65	0.92	1.40	0.2	3.4	21,107
Panel C: *Down Markets*										
CSAD	0.46	0.30	0.01	0.18	0.38	1.02	4.32	2.6	15.0	19,645
Market Return	−0.63	0.78	−10.81	−2.12	−0.38	−0.03	−0.00	−3.3	21.0	19,645
Trading Volume	8.00	14.04	0.00	0.55	3.18	31.23	333.88	6.1	71.4	19,645
Blockchain Transactions	3.01	0.42	1.80	2.39	2.94	3.72	4.17	0.3	2.1	19,645
Search Volume$_{\text{Level}}$	1.98	0.69	0.47	1.09	1.86	3.32	4.51	0.8	3.3	19,645
Search Volume$_{\text{Dispersion}}$	0.82	0.22	0.10	0.42	0.84	1.17	2.13	−0.2	3.3	19,645
Reddit Posts$_{\text{Level}}$	2.13	0.59	0.40	1.23	2.11	3.14	4.46	0.3	2.8	19,645
Reddit Posts$_{\text{Dispersion}}$	0.66	0.16	0.11	0.40	0.65	0.92	1.55	0.2	3.4	19,645

This table shows summary statistics for our key variables. *CSAD* is the cross-sectional absolute deviation of returns in percent. *Market Return* is the hourly logarithmic return of the market index in percent. *Squared Market Return* is the squared hourly logarithmic return of the market index in basis points. *Blockchain Transactions* is the equally weighted cross-sectional average of the log normalized number of transactions recorded on the blockchain within an hour. $Search\ Volume_{Level}$ is the equally weighted cross-sectional average of the log normalized Google search volume within an hour. Similarly, $Search\ Volume_{Dispersion}$ is its cross-sectional absolute deviation. $Reddit_{Level}$ and $Reddit_{Dispersion}$ are constructed analogously using the number of submissions and comments on Reddit. *Trading Volume* is the total hourly trading volume of all currencies in the market in 1mn USD. In Panel A, the full sample is used. In Panels B and C, the sample is split into observations with positive and negative market returns, respectively.

Table 4.2: Correlations

	CSAD	Market Return	Market Return^2	Trading Volume	Blockchain Trans.	Search Volume$_{\text{Level}}$	Search Volume$_{\text{Disp.}}$	Reddit Posts$_{\text{Level}}$
Market Return	0.01							
Market Return^2	0.35	−0.18						
Trading Volume	0.27	−0.07	0.33					
Blockchain Transactions	0.05	−0.00	0.07	0.51				
Search Volume$_{\text{Level}}$	0.38	−0.00	0.15	0.43	0.48			
Search Volume$_{\text{Dispersion}}$	−0.15	0.00	−0.03	0.01	0.12	−0.40		
Reddit Posts$_{\text{Level}}$	0.37	−0.01	0.13	0.25	0.17	0.73	−0.29	
Reddit Posts$_{\text{Dispersion}}$	0.04	0.00	0.02	0.12	0.18	0.08	0.07	0.11

This table shows correlations across the full sample for the variables as defined in Table 4.1.

Table 4.3: Baseline Herding Analysis

	(1)	(2)	(3)	(4)	(5)	(6)
|Market Return|	0.242***	0.225***	0.227***	0.158***	0.146***	0.144***
	(36.50)	(33.30)	(33.37)	(26.16)	(24.62)	(24.31)
Market Return^2	−1.238***	−1.260***	−1.377***	−0.457**	−0.608***	−0.589***
	(−6.59)	(−6.84)	(−7.29)	(−2.42)	(−3.34)	(−3.21)
Trading Vol.		0.002***	0.003***		0.004***	0.003***
		(7.93)	(9.81)		(12.41)	(11.74)
Blockchain Trans.			−0.059***			0.108***
			(−5.96)			(7.08)
Date FE	–	–	–	✓	✓	✓
Observations	40799	40799	40799	40798	40798	40798
Adj.R^2	0.212	0.222	0.227	0.533	0.543	0.544

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. The variables are as defined in Table 4.1, except CSAD which is here given in basis points. T-statistics with Newey and West (1987) standard errors are reported in parentheses, except for in the fixed effects model, where the standard errors are clustered by date. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

4.4.2 Baseline Herding Analysis

We now investigate herding behavior and its potential determinants. The baseline results can be found in Table 4.3. In the first model we apply the basic specification of Chang et al. (2000) to our hourly data. We find a significantly negative coefficient for the squared market returns, suggesting that investors exhibit return herding behavior and confirming our first hypothesis. This result thus agrees with several previous studies that find herding in cryptocurrency markets (e.g. Kaiser and Stöckl, 2020; Ballis and Drakos, 2020), though the size and significance of the effect appear to support the notion of Gleason et al. (2004) that higher frequency returns may be better able to detect short-term herding by investors.

We then include the trading volume at Kraken and the number of transactions recorded on the blockchains. For the exchange-specific trading volume we find a positive and significant coefficient. This means that the higher the aggregate demand to exchange these specific currencies (in particular against the USD), the less investors herd around the market, consistent with the finding by Youssef (2022b), who find that investors herd less as the trading volume increases. For the number of transactions, we find a negative coefficient. This indicates that over the course of the sample period, a higher demand in blockchain transactions is associated with more herding, though neither the adjusted R^2 nor the size of the return herding coefficient meaningfully change after including the transaction count. On a longer horizon, the transaction count correlates with the popularity of a currency and how broad the investor base is. Taken together, the negative coefficient would thus indicate that the more popular the currencies in the sample get, the more investors herd around the market.

The hourly data allows us to include date fixed effects to control for overall trends in the data. The identification of herding behavior and how it relates to market returns and the other investigated potential determinant of herding now comes from their intraday variation. Including date fixed effects thus focuses the analysis on short-term herding, whereas the overall analysis before includes both short-term and longer-term effects. The results are presented in the rightmost three columns of Table 4.3. Overall, our conclusion of significant market return herding behavior proves robust in all models. In the baseline model, we find that the size of the effect of squared market returns on cross-sectional dispersion reduces to about one third but stays highly statistically significant. The biggest difference from including date fixed effects can be found in the effect of the number of transactions. While in the analysis without date-fixed effects, more transactions are associated with less return dispersion, the opposite is true when focusing on intraday variations by including these fixed effects. This suggests that the short and long run effects of transaction activity go in opposite directions. In the long run, the measure likely reflects the currencies' popularity and broadness of investor base, while in the short run the measure is more likely to capture the activity of roughly the same investor base.

4.4.3 Herding and Investor Attention

We then study how investor attention relates to herding. The results are presented in Panel A of Table 4.4. First, we include the level of internet search volume as a proxy for information demand by investors and find a significantly positive relationship with the dispersion of returns. In other words, a higher level of investor attention as measured by the aggregate search activity across different cryptocurrencies is associated with lower levels of herding around the market consensus, contrary to the results found for stock markets by Hsieh et al. (2020). The magnitude of the return herding coefficient is reduced by about 25% while the adjusted R^2 increases slightly, indicating that investor information demand is indeed an important determinant of investor herding. Similarly, we then include search volume dispersion and find a positive effect on return dispersion. In fact, the effect of the level of attention is stronger when dispersion is additionally included, which implies that they capture two different dimensions of attention. This is further verified by the time-series correlation of about -40% between the level and dispersion of search volume.[10] Interpreting these results jointly, we find evidence that the more investors search for cryptocurrency information, the more idiosyncratic information they incorporate into their trading decisions and thus prices. Likewise, the more dispersed their searches are across the individual currencies, the less their trading decisions reflect the market consensus since their searches are less likely to solely reflect an increase in general interest in cryptocurrencies, leading to less herding. Overall, we thus find strong support for hypothesis 4.

We then turn to the investor attention measures capturing the supply of information, where we proceed similarly as with search volume. In models 3–4, we include the level and dispersion of the number of posts on the message boards for the different cryptocurrencies on the platform Reddit. The results generally mimic those found for search volume, though they are smaller in magnitude. The level of attention is significantly negatively associated with herding around the market return. However, in this specification, the effect of dispersion is insignificant. Note that the results regarding Reddit posts do not necessarily imply that investors receive correct fundamental information through this channel. As pointed out by Sabherwal et al. (2011), pump-and-dump schemes facilitated by message boards might lead investors to drive prices of individual currencies further away from the market, thus increasing the cross-sectional dispersion of returns.

In model 5, we include both groups of attention measures. While the magnitudes of the coefficients for search volume and the level of Reddit posts decrease, the effect of dispersion in Reddit posts now turns significant. This suggests that search volume and message board posts indeed capture different dimensions of investor attention and that both are negatively related to herding around the market portfolio, further confirming hypothesis 4.

In Panel B we repeat the analysis by including date fixed effects so that the identification comes from the intraday variation in the data. Again, all results prove robust to controlling for long-term trends. Contrary to before, the coefficient for squared market returns now is hardly affected by the inclusion of the different attention measures. Interpreting these results jointly, this suggests that long-term changes in market consensus herding behavior (as defined in Chang et al., 2000) are partially explained by long-term changes in aggregate attention, but only to a lesser extent by short-term fluctuations in attention.

Generally speaking, the coefficients for the level of attention decrease in magnitude while those for attention dispersion increase compared to the results without date fixed effects, in particular when including all four variables in model 5. Overall, these findings suggest that intraday variations in attention are strongly related to conditional herding behavior (as defined in Bernales et al., 2020), which will be investigated in more detail below.

Table 4.4: Herding and Investor Attention

	(1)	(2)	(3)	(4)	(5)
Panel A: *No Fixed Effects*					
|Market Return|	0.194***	0.192***	0.198***	0.198***	0.190***
	(30.44)	(30.44)	(30.26)	(30.25)	(29.70)
Market Return^2	−0.937***	−0.912***	−1.021***	−1.021***	−0.891***
	(−5.11)	(−5.02)	(−5.46)	(−5.46)	(−4.85)
Trading Vol.	0.002***	0.002***	0.002***	0.002***	0.002***
	(5.41)	(5.10)	(6.91)	(6.90)	(5.09)
Blockchain Trans.	−0.155***	−0.170***	−0.072***	−0.072***	−0.149***
	(−18.17)	(−17.83)	(−8.37)	(−8.19)	(−15.33)
Search Vol.$_{\text{Level}}$	0.160***	0.176***			0.131***
	(24.04)	(23.81)			(18.82)
Search Vol.$_{\text{Dispersion}}$		0.077***			0.058***
		(6.13)			(4.67)
Reddit Posts$_{\text{Level}}$			0.151***	0.151***	0.058***
			(20.54)	(20.50)	(8.72)
Reddit Posts$_{\text{Dispersion}}$				0.002	0.033**
				(0.11)	(2.09)
Date FE	–	–	–	–	–
Observations	40799	40799	40799	40799	40799
Adj.R^2	0.318	0.320	0.303	0.303	0.326
Panel B: *Date Fixed Effects*					
|Market Return|	0.144***	0.144***	0.144***	0.144***	0.144***
	(24.34)	(24.38)	(24.19)	(24.21)	(24.28)
Market Return^2	−0.587***	−0.590***	−0.581***	−0.578***	−0.579***
	(−3.19)	(−3.22)	(−3.14)	(−3.13)	(−3.15)
Trading Vol.	0.003***	0.003***	0.003***	0.003***	0.003***
	(11.54)	(11.38)	(11.25)	(11.15)	(10.81)
Blockchain Trans.	0.104***	0.118***	0.077***	0.070***	0.083***
	(6.90)	(7.70)	(4.97)	(4.57)	(5.35)
Search Vol.$_{\text{Level}}$	0.024***	0.052***			0.049***
	(5.69)	(8.11)			(7.80)
Search Vol.$_{\text{Dispersion}}$		0.063***			0.068***
		(6.16)			(6.64)
Reddit Posts$_{\text{Level}}$			0.046***	0.045***	0.044***
			(7.99)	(7.65)	(7.67)
Reddit Posts$_{\text{Dispersion}}$				0.057***	0.057***
				(5.73)	(5.73)
Date FE	✓	✓	✓	✓	✓
Observations	40798	40798	40798	40798	40798
Adj.R^2	0.545	0.545	0.545	0.546	0.547

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panel A does not include fixed effect, whereas Panel B includes day fixed effects. The variables are as defined in Table 4.1, except CSAD which is here given in basis points. T-statistics with Newey and West (1987) standard errors are reported in parentheses, except for in the fixed effects model, where the standard errors are clustered by date. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.5: Herding in Up and Down Markets

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Up Markets*						
|Market Return|	0.290***	0.266***	0.229***	0.203***	0.187***	0.187***
	(30.92)	(27.33)	(25.47)	(24.51)	(22.20)	(22.40)
Market Return^2	−2.031***	−1.955***	−1.515***	−1.143***	−1.151***	−1.151***
	(−6.12)	(−5.79)	(−4.64)	(−3.24)	(−3.26)	(−3.30)
Trading Vol.		0.004***	0.002***		0.004***	0.004***
		(7.73)	(4.22)		(8.95)	(8.29)
Blockchain Trans.		−0.069***	−0.159***		0.125***	0.100***
		(−5.27)	(−12.52)		(6.44)	(5.05)
Search Vol.$_{\text{Level}}$			0.135***			0.056***
			(15.06)			(5.88)
Search Vol.$_{\text{Dispersion}}$			0.043***			0.074***
			(2.69)			(4.75)
Reddit Posts$_{\text{Level}}$			0.055***			0.041***
			(6.28)			(5.11)
Reddit Posts$_{\text{Dispersion}}$			0.040*			0.068***
			(1.92)			(4.71)
Date FE	–	–	–	✓	✓	✓
Observations	21107	21107	21107	21104	21104	21104
Adj.R^2	0.210	0.228	0.325	0.521	0.534	0.538
Panel B: *Down Markets*						
|Market Return|	0.197***	0.188***	0.150***	0.114***	0.103***	0.104***
	(24.15)	(21.78)	(18.87)	(15.53)	(13.97)	(13.91)
Market Return^2	−0.470**	−0.677***	−0.168	0.261	0.049	0.061
	(−2.29)	(−3.18)	(−0.83)	(1.28)	(0.24)	(0.29)
Trading Vol.		0.003***	0.001***		0.003***	0.003***
		(7.32)	(3.27)		(11.25)	(10.48)
Blockchain Trans.		−0.051***	−0.138***		0.084***	0.061***
		(−4.07)	(−11.58)		(4.36)	(3.11)
Search Vol.$_{\text{Level}}$			0.128***			0.041***
			(15.59)			(6.05)
Search Vol.$_{\text{Dispersion}}$			0.070***			0.062***
			(4.42)			(4.98)
Reddit Posts$_{\text{Level}}$			0.063***			0.043***
			(8.32)			(6.19)
Reddit Posts$_{\text{Dispersion}}$			0.021			0.047***
			(1.12)			(3.64)
Date FE	–	–	–	✓	✓	✓
Observations	19645	19645	19645	19643	19643	19643
Adj.R^2	0.226	0.237	0.343	0.560	0.570	0.573

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panels A and B only include observations during positive or negative market returns, respectively. The variables are as defined in Table 4.1, except CSAD which is here given in basis points. T-statistics with Newey and West (1987) standard errors are reported in parentheses, except for in the fixed effects model, where the standard errors are clustered by date. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

4.4.4 Herding across different Market States

Because several prior studies have proposed asymmetries in herding behavior in various markets (Chang et al., 2000; Chiang and Zheng, 2010; Vidal-Tomás et al., 2019), we proceed

[10] See Table 4.2 in the appendix for all time-series correlations.

by splitting the sample into periods of positive and negative market returns. The results in Table 4.5 show that investors particularly herd in bull market states. For down markets, the herding coefficients are negative and significant when not controlling for attention or date fixed effects, but otherwise insignificant. Our results thus agree with Ballis and Drakos (2020), who also find some herding behavior by cryptocurrency investors in both up and down markets, but a stronger effect when prices are increasing. When compared to non-cryptocurrency markets, our results resemble those found for many Asian equity markets in Chiang and Zheng (2010): Herding exists in both market states, but there is an overall stronger effect during up markets. This behavior is consistent with the idea that cryptocurrency investors are prone to trading based on a fear-of-missing-out and short sell restriction. Observing that market prices are increasing, the average investor does not want to miss out on bullish markets and hence similarly invests across the cryptocurrency universe; on the other hand, the short-sell restriction limits herding when the market prices are decreasing. This effect is potentially aggravated by short-sell restrictions so that investors with opposing views may find it difficult to trade on their information. We thus find support for hypothesis 2a and reject hypothesis 2b.

The other estimated coefficients stay at similar levels and significances compared to the unconditional model. In particular, this suggests that the effects of investor attention on herding do not materially depend on the market state. Likewise, in the descriptive statistics we found virtually identical levels and dispersions of attention for up and down markets. The differences in return herding behavior between up and down markets are thus unlikely to be driven by differences in attention, but rather indicate that investor attention and the direction of market returns are two distinct determinants of investor herding.

Furthermore and similarly to Kallinterakis and Wang (2019), we split the sample into high and low volatility periods. High market volatility periods are defined as those where the estimated volatility is larger than its moving average of the previous two weeks, where we estimate volatility for every hour using the forward-looking asymmetric power ARCH model of Ding et al. (1993). The results are presented in Table 4.6 in the appendix and show that herding is more prominent during low volatility periods, agreeing with the result of Kallinterakis and Wang (2019), although contrary to the finding by Raimundo Júnior et al. (2022), Youssef (2022b) and Youssef and Waked (2022). Nonetheless, we find significant herding in some specifications during high volatility periods as well.

Finally, Table 4.7 in the appendix shows that there are some differences in herding between weekdays and weekends. In particular, herding behavior appears to be much stronger during the weekend. Assuming that the fraction of small retail traders is larger during the weekend, this result is consistent with the notion that these investors are particularly prone to exhibit herd behavior.[11]

[11]Untabulated t-tests show that average trade sizes are smaller during the weekend, which is consistent with the assumption of a larger fraction of retail traders. See also Scharnowski (2021b).

4.4.5 Intraday Patterns in Herding and Attention

So far, the higher granularity of our dataset has allowed us to document significant market return herding behavior while controlling for general trends in the data. In this part, we analyze how investor attention and herding behavior fluctuate throughout the day.

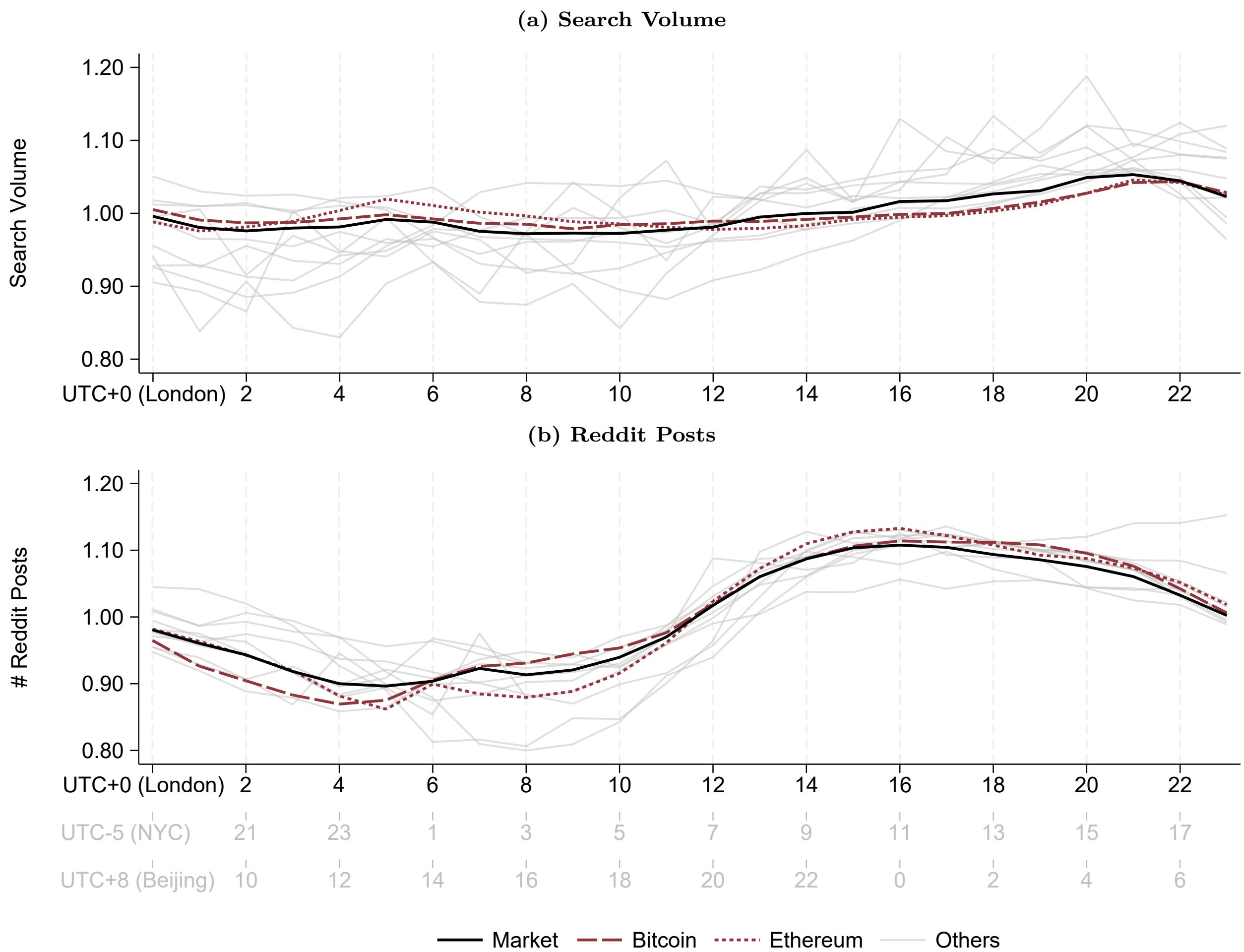
All timestamps used in this study are in Coordinated Universal Time (UTC), though this is simply a convention. Cryptocurrencies trade around the clock in a mostly anonymous, decentralized way, which makes it difficult to know where traders are located geographically. While we only consider trading against the USD, this does not necessarily imply that those traders are located in the United States. Even if they were, it is ex ante unclear when during the day traders would trade: Professional investors likely trade during regular business hours, retail investors might exhibit different patterns of daily trading activity, whereas algorithms trade throughout the day. When visualizing the intraday patterns, we hence provide the times of two additional time zones. Disregarding daylight-saving time, these roughly correspond to the time in New York City and Beijing, respectively.

To provide a ballpark approximation of how the potential non-algorithmic investor base fluctuates throughout the day, we estimate the worldwide population with internet access that is awake at a given point in time from the perspective of an investor in UTC+0. We combine data from the International Telecommunication Union and the United Nations World Population Prospects and further assume that for a given time zone, half the population is awake between 6:00h and 8:00h, the full population is awake between 8:00h and 23:00h, and again half the population is awake between 23:00h and 01:00h (in local time). For countries spanning multiple time zones, the geographical distribution of internet users is assumed to be identical to that of the overall population. Under these assumptions, we find that the potential investor base is largest from 11:00h to 14:00h UTC+0, which would be the morning in eastern North and all of South America, mid-day in Europe and Africa and the evening in large parts of Asia. The finding coincides with the one of Brauneis et al. (2023) who find that trading activity is lowest in the early morning UTC hours and highest around 15:00 and 16:00 UTC. The intraday variation of the online population can be found in Figure 4.3 in the appendix.

Before analyzing intraday herding in detail, we turn to the intraday patterns in investor attention. Figure 4.1a shows how attention as measured by the level of search volume evolves throughout the day. We make several observations: First, each individual currency exhibits intraday variation in search volume. The largest currencies show less variation than the smaller ones. Second, there is some co-movement of attention. For example, search volume is generally higher at 21:00 UTC than at 08:00 UTC. The average attention level across the currencies is bimodal with peaks at 05:00 UTC and 21:00 UTC. Third, the co-movement is less than perfect, leading to fluctuations in intraday attention dispersion.[12]

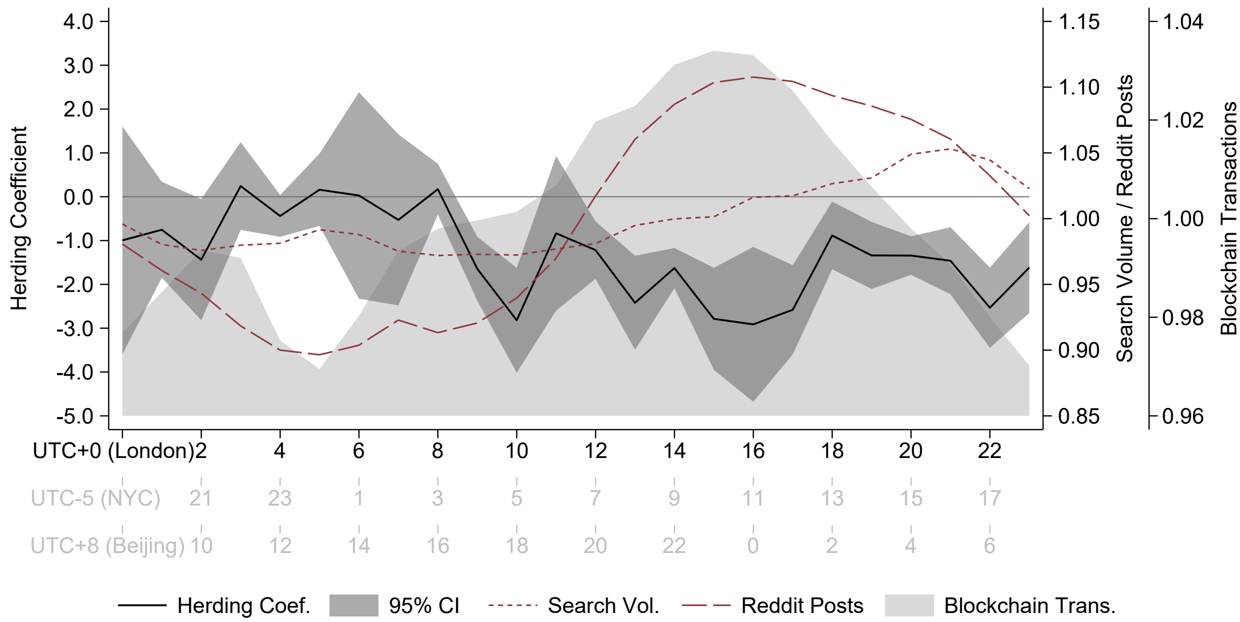
[12] Note that for this graph, the individual time series are scaled by their means to visualize the co-movement. The downside is that attention dispersion is not directly visible in the graph because it depends on the level of differences. The corresponding graph of search volume attention dispersion can be found in Figure 4.4 in the appendix.

Figure 4.1: Intraday Patterns in Investor Attention



These graphs show intraday variations in investor attention. The first graph shows intraday variations in investor attention as measured by Google search volume. The bottom graph shows intraday variations in information supply as measured by Reddit submissions and comments. The data is normalized and averaged for every one hour windows of the day. Bitcoin and Ethereum have been highlighted. *Market* is the equally weighted average across all currencies. For the purpose of this graph, all time series have been standardized by dividing by their respective averages.

Figure 4.2: Intraday Market Return Herding Patterns



This graph shows the regression coefficients $\beta_{2,h}$ from estimating Equation 4.2. The dark gray area indicate the 95% confidence interval for the coefficient estimate based on Newey and West (1987) standard errors. Significantly negative coefficients indicate herding. The dotted line shows the average level of search volume across the sample currencies. The long-dashed line shows the average number of posts on Reddit. Additionally, the light gray area shows the average number of transactions on the blockchains of the sample currencies. Search volume, Reddit posts, and blockchain transactions have been standardized by dividing by the respective overall means of the two time series.

Similarly, Figure 4.1b shows the intraday development of the scaled number of Reddit posts which measures the information supply aspect of investor attention. Again, all individual currencies exhibit some form of intraday variation. The number of posts is generally higher during the second half of the (UTC+0) day, when investors in Europe and the Americas are likely awake. This probably reflects the geographic distribution of the user base, as Reddit is relatively more popular in the United States than in other parts of the world. The number of new posts for larger cryptocurrencies tends to fluctuate less than for the smaller ones, again leading to differences in the intraday dispersion of attention throughout the day.

We then estimate intraday return herding patterns by Equation 4.2. The results in Figure 4.2 show the coefficients of squared market returns for every hour of the day, where significantly negative coefficients indicate market return herding. We find a distinct pattern in intraday herding activity: From 00:00 to about 08:00 (all in UTC+0), which corresponds to nighttime in Europe and large parts of the Americas, the herding coefficient is mostly insignificant. Roughly from 10:00 to 17:00, we observe the largest absolute values for the herding coefficient, which, except for the value at 11:00, are all statistically significant. These times also contain the largest overlap of potential investors likely being awake and the overlap of conventional exchange trading hours in Europe and North America. Herding then decreases for the rest of the day but stays significant at the 5% level. Overall, these results are consistent with hypothesis 5.

The figure also shows the normalized aggregate measures of attention and the normalized aggregate number of transactions recorded on the blockchain. As expected, their corresponding graphs are similar: Intraday periods of high herding activity closely coincide with periods where many transactions are recorded on the blockchains. Furthermore, there is a positive relation between attention and intraday herding around the market consensus.

There is an important difference between this analysis and the previous regressions. In the preceding section, we relate the cross-sectional dispersion of returns to investor attention and activity in the spirit of Bernales et al. (2020). The results indicate that attention and investor activity are associated with higher return dispersion and hence less "conditional herding". This contrasts with the analysis presented here, where we relate the market return herding coefficient, obtained from regressing CSAD on squared market returns, to attention and trading activity. We here thus explicitly capture how the non-linear reaction of return dispersion to more extreme market returns is affected by these other variables. In other words, while previously we found that higher levels of attention and activity are associated with less herding across the entire return distribution, here we find that these variables are positively associated with herding during times of large market movements.[13]

Furthermore, the documented intraday patterns suggest that trading in cryptocurrencies is not fully automated but instead are consistent with a material role of deliberate trading decisions by retail and possibly institutional investors. Our findings thus agree with and supplement those found in Petukhina et al. (2021) and Baur et al. (2019).

4.4.6 Robustness

We perform several robustness tests. We first confirm in untabulated estimations that additionally including the signed market return in the regressions does not meaningfully impact our results. The same holds for including a market volatility proxy which we estimate for every hour using the asymmetric power ARCH model.

Moreover, untabulated results show that our findings are generally robust to using different market indices. Firstly, we construct a value weighted market index using the square root of market capitalization, thus putting more weight on larger cryptocurrencies. We then employ the extreme case of only using Bitcoin returns as the market index, acknowledging that Bitcoin is often used as a transfer currency (Kaiser and Stöckl, 2020). While these other weighting schemes generally lead to lower estimates of return herding behavior, we still find similar intraday patterns, which are shown in Figure 4.6 in the appendix.

Since we document strong intraday patterns in herding and attention, a natural question might be whether our results are driven by some other, omitted factor with similar intraday patterns. First note that in another robustness test above we already control for volatility and thus its known intraday patterns (Petukhina et al., 2021). Still, we address this concern

[13]For reference, Figure 4.5 shows the development of CSAD throughout the day. Note that while there are some similarities, clearly the herding coefficient exhibits different intraday variations than return dispersion.

by including fixed effects for every hour of the day. Table 4.8 in the appendix shows that our results are indeed robust to including such intraday fixed effects.

Our results are also robust to inclusion of further control variables. Following the approach by Christie and Huang (1995), the estimation results when monthly variables, such as the monthly average of the hourly market return, are included as control variables in Table 4.9 for baseline specification in Table 4.3 and in Table 4.10 for specifications involving all variables intoduced in Table 4.1. In Table 4.11, the monthly returns of stock indices from different continents are included as control variables. In all instances, the main results persist.[14] The asymmetries in herding behavior also persists after including further control variables, as shown in Table 4.12 for up and down market and Table 4.13 for market during high and low volatility.

Similar observations can be made regarding intraday patterns in herding behavior. The results in Figure 4.7 show the herding coefficients for every hour of the day, which, in contrast to Figure 4.2, are estimated controlling for these variables: *Trading Volume*, *Blockchain Transactions*, $Search\ Volume_{Level}$, $Search\ Volume_{Dispersion}$, $Reddit_{Level}$, and $Reddit_{Dispersion}$. While the herding coefficients from 10:00 to 17:00 in Figure 4.7 exhibit less significance than the ones in Figure 4.2, the overall pattern remains consistent.

4.5 Conclusion

Using a rich dataset of intraday return dispersion, attention, and transaction activity, we document the presence of substantial market return herding behavior in the cryptocurrency market. Both the level of investor attention, as measured by internet searches and message board posts, and its dispersion increase the cross-sectional dispersion of returns and thus lead to reduced herding behavior. The more investors search for (or are confronted with) coin-specific information, the more idiosyncratic information they incorporate into their trading decisions and thus prices. Likewise, the less investors encounter the same information, the less their individual trading decisions reflect the market consensus. Higher transactional activity on the currencies' blockchains has mixed effects on herding. In the short run, it is associated with decreased herding behavior. However, in the long run where transactional activity likely correlates with the popularity of the currencies and the breadth of their investor base, the effect turns around so that more transactions are associated with more herding activity. Additionally, we find that investors follow the market more closely during bull markets.

Zooming into potential intraday patterns, we find that herding varies substantially throughout the day. It is strongest during the overlap of hours when traders in major economic centers are likely awake. At the intraday level, investor information demand and supply, blockchain transaction activity, and exchange trading volume are positively correlated with

[14] It is interesting to note that the coefficients of stock indices from the same continents consistently maintain both their significance and sign. Despite this, the sign, significance, and magnitude of the coefficients remain remarkably consistent.

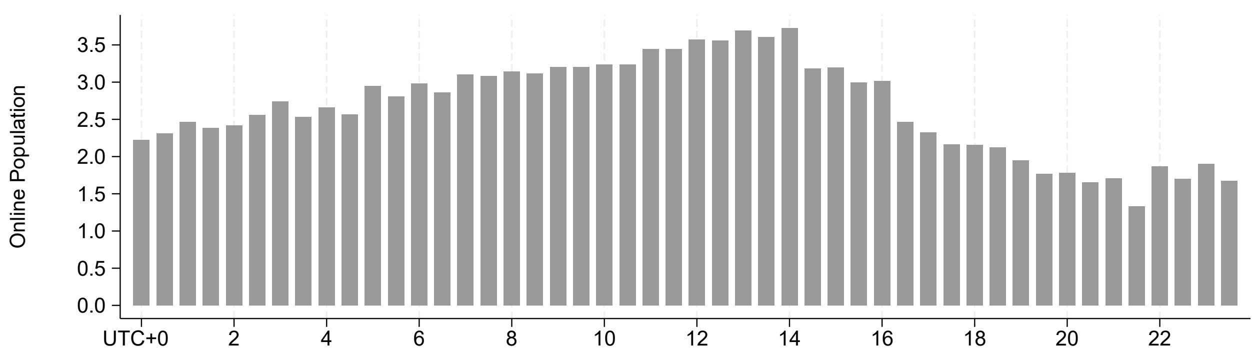
the extent of herding behavior. These results are consistent with the presence of retail or unsophisticated institutional investors.

Our results have important implications. Market return herding might deteriorate market quality and lead to inefficient prices as investors disregard the already scarce fundamental information, potentially creating irrational bubbles. Understanding how herding, trading activity, and investor attention are related thus helps traders and regulators to design better trading strategies and more resilient markets, taking into account the particularities of each market. For example, in the case of the cryptocurrency market where little fundamental information is available, educating potential investors about the assets might lead to less herding, since we document that investors are generally willing to search for specific information and incorporate them into prices instead of always blindly following the market. However, they do not consistently choose to gather idiosyncratic information, so there is room for regulatory improvement assuming such information is generally available.

While in this paper we shed light on some of the determinants of intraday herding in cryptocurrency markets, further research may investigate additional potential determinants. Similar to studies on other financial markets, these could include changes in informational supply, regulatory interventions, or spillovers from other markets.

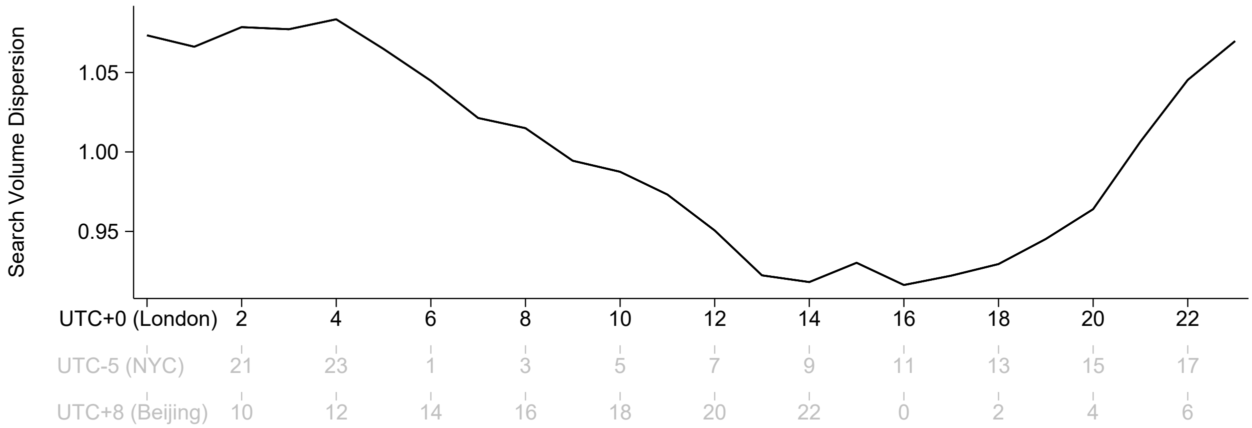
4.6 Appendix to Chapter 4

Figure 4.3: Online Population throughout the Day



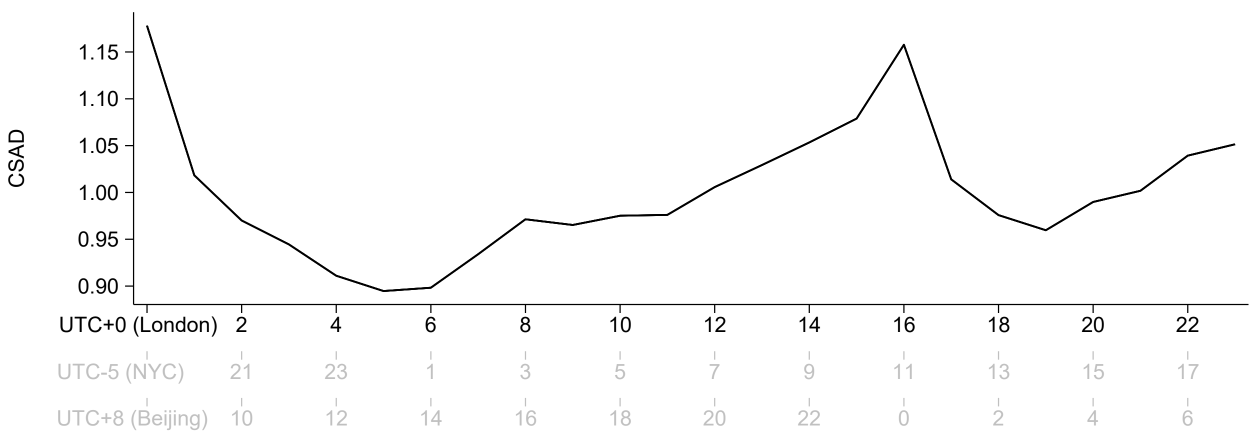
This graph shows an estimate of the worldwide population (in 1bn) with internet access that is awake during each 30 minute window from the perspective of UTC+0. The values are estimated by combining the percentage of the population with internet access (from the International Telecommunication Union) with the total population (from the United Nations World Population Prospects) for every country. The data is then aggregated to time zones while ignoring daylight saving time. For countries spanning multiple time zones, the geographical distribution of internet users is assumed to be identical to that of the overall population. It is further assumed that for a given time zone, half the population is awake between 6:00h and 8:00h, the full population is awake between 8:00h and 23:00h, and again half the population is awake between 23:00h and 01:00h (in local time).

Figure 4.4: Intraday Patterns in Search Volume Dispersion



This graph shows hourly averages of investor attention dispersion as measured by the cross-sectional absolute deviation of the log normalized Google search volume from the market average level of attention. For the purpose of this graph, the time series has been standardized by dividing by its mean.

Figure 4.5: Intraday Patterns in CSAD



This graph shows hourly averages of the cross-sectional absolute deviation of returns from the market return. For the purpose of this graph, the time series has been standardized by dividing by its mean.

Table 4.6: Herding during High and Low Market Volatility

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *High Market Volatility*						
|Market Return|	0.218***	0.206***	0.170***	0.144***	0.133***	0.133***
	(24.34)	(23.30)	(19.77)	(16.91)	(16.01)	(15.91)
Market Return^2	−0.788***	−0.906***	−0.516**	−0.215	−0.355	−0.351
	(−3.39)	(−4.09)	(−2.28)	(−0.91)	(−1.55)	(−1.53)
Trading Vol.		0.003***	0.002***		0.004***	0.003***
		(7.17)	(3.90)		(9.00)	(8.48)
Blockchain Trans.		−0.091***	−0.194***		0.105***	0.085***
		(−5.59)	(−12.23)		(3.92)	(3.15)
Search Vol.$_{\text{Level}}$			0.146***			0.063***
			(13.52)			(4.72)
Search Vol.$_{\text{Dispersion}}$			0.056***			0.078***
			(2.80)			(3.86)
Reddit Posts$_{\text{Level}}$			0.050***			0.044***
			(4.49)			(4.03)
Reddit Posts$_{\text{Dispersion}}$			0.045			0.062***
			(1.63)			(3.25)
Date FE	–	–	–	✓	✓	✓
Observations	16365	16365	16365	16296	16296	16296
Adj.R^2	0.211	0.227	0.316	0.510	0.524	0.527
Panel B: *Low Market Volatility*						
|Market Return|	0.247***	0.230***	0.194***	0.177***	0.166***	0.166***
	(28.38)	(25.48)	(23.33)	(28.27)	(27.06)	(27.41)
Market Return^2	−2.045***	−2.423***	−1.602***	−1.027***	−1.297***	−1.295***
	(−7.07)	(−8.11)	(−5.89)	(−5.36)	(−6.78)	(−6.88)
Trading Vol.		0.004***	0.002***		0.003***	0.003***
		(8.07)	(3.83)		(10.37)	(9.32)
Blockchain Trans.		−0.050***	−0.124***		0.115***	0.089***
		(−4.62)	(−11.81)		(6.61)	(5.03)
Search Vol.$_{\text{Level}}$			0.118***			0.033***
			(14.52)			(5.53)
Search Vol.$_{\text{Dispersion}}$			0.057***			0.060***
			(4.08)			(5.44)
Reddit Posts$_{\text{Level}}$			0.060***			0.045***
			(8.71)			(6.90)
Reddit Posts$_{\text{Dispersion}}$			0.019			0.051***
			(1.19)			(4.88)
Date FE	–	–	–	✓	✓	✓
Observations	24434	24434	24434	24387	24387	24387
Adj.R^2	0.165	0.182	0.296	0.520	0.529	0.532

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panels A and B only include observations during high and low market return volatility, respectively. Volatility is estimated by an asymmetric power ARCH model. High market volatility periods are defined as those where the estimated volatility is larger than its moving average of the previous two weeks. The variables are as defined in Table 4.1, except CSAD which is here given in basis points. T-statistics with Newey and West (1987) standard errors are reported in parentheses, except for in the fixed effects model, where the standard errors are clustered by date. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.7: Herding during the Weekend

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *During the Week*						
|Market Return|	0.227***	0.212***	0.181***	0.151***	0.136***	0.136***
	(28.38)	(26.13)	(23.24)	(19.76)	(18.11)	(18.09)
Market Return^2	−0.908***	−1.020***	−0.672***	−0.303	−0.387	−0.383
	(−3.77)	(−4.20)	(−2.76)	(−1.18)	(−1.55)	(−1.53)
Trading Vol.		0.003***	0.002***		0.004***	0.003***
		(7.52)	(3.91)		(10.45)	(9.61)
Blockchain Trans.		−0.060***	−0.151***		0.107***	0.081***
		(−4.98)	(−13.60)		(5.98)	(4.43)
Search Vol.$_{\text{Level}}$			0.134***			0.047***
			(15.95)			(6.06)
Search Vol.$_{\text{Dispersion}}$			0.061***			0.066***
			(4.17)			(5.26)
Reddit Posts$_{\text{Level}}$			0.057***			0.044***
			(7.46)			(6.63)
Reddit Posts$_{\text{Dispersion}}$			0.040**			0.055***
			(2.19)			(4.51)
Date FE	–	–	–	✓	✓	✓
Observations	29066	29066	29066	29065	29065	29065
Adj.R^2	0.205	0.219	0.317	0.527	0.539	0.542
Panel B: *During the Weekend*						
|Market Return|	0.277***	0.260***	0.211***	0.174***	0.165***	0.165***
	(24.68)	(21.75)	(19.36)	(17.59)	(16.99)	(16.98)
Market Return^2	−1.951***	−2.199***	−1.402***	−0.804***	−1.038***	−1.014***
	(−6.82)	(−7.29)	(−5.17)	(−3.43)	(−4.72)	(−4.63)
Trading Vol.		0.005***	0.002***		0.003***	0.003***
		(8.72)	(4.52)		(5.57)	(5.23)
Blockchain Trans.		−0.062***	−0.149***		0.109***	0.086***
		(−4.41)	(−8.94)		(3.78)	(2.99)
Search Vol.$_{\text{Level}}$			0.129***			0.052***
			(11.38)			(5.03)
Search Vol.$_{\text{Dispersion}}$			0.060***			0.072***
			(2.78)			(4.07)
Reddit Posts$_{\text{Level}}$			0.053***			0.044***
			(4.46)			(3.85)
Reddit Posts$_{\text{Dispersion}}$			0.013			0.059***
			(0.47)			(3.73)
Date FE	–	–	–	✓	✓	✓
Observations	11733	11733	11733	11733	11733	11733
Adj.R^2	0.229	0.250	0.347	0.546	0.554	0.558

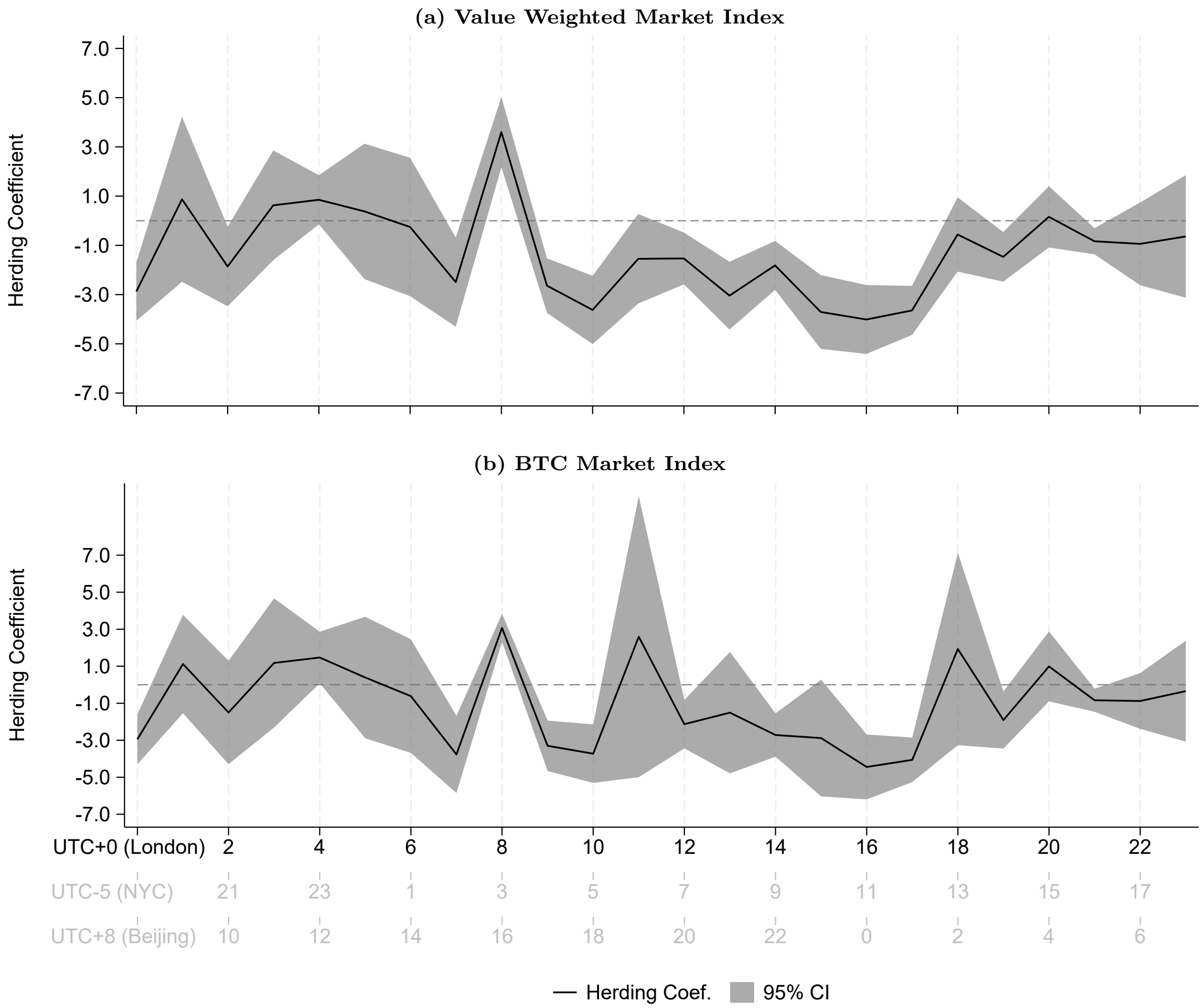
This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panels A and B only include observations during the week and the weekend, respectively. The weekend is defined as Saturday and Sunday UTC+0. The variables are as defined in Table 4.1, except CSAD which is here given in basis points. T-statistics with Newey and West (1987) standard errors are reported in parentheses, except for in the fixed effects model, where the standard errors are clustered by date. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.8: Herding during all Market States with Intraday FE

	(1)	(2)	(3)	(4)	(5)
Panel A: *Intraday Fixed Effects*					
|Market Return|	0.189***	0.187***	0.194***	0.194***	0.186***
	(27.44)	(27.41)	(27.42)	(27.43)	(26.97)
Market Return^2	−0.878***	−0.837***	−0.971***	−0.971***	−0.832***
	(−4.78)	(−4.63)	(−5.26)	(−5.26)	(−4.57)
Trading Vol.	0.002***	0.002***	0.002***	0.002***	0.002***
	(4.76)	(4.33)	(6.31)	(6.29)	(4.50)
Blockchain Trans.	−0.161***	−0.185***	−0.070***	−0.071***	−0.163***
	(−15.92)	(−15.72)	(−6.98)	(−6.93)	(−13.19)
Search Vol.$_{\text{Level}}$	0.164***	0.186***			0.147***
	(21.68)	(21.37)			(16.82)
Search Vol.$_{\text{Dispersion}}$		0.112***			0.081***
		(7.29)			(5.11)
Reddit Posts$_{\text{Level}}$			0.158***	0.158***	0.046***
			(18.05)	(18.03)	(5.46)
Reddit Posts$_{\text{Dispersion}}$				0.017	0.027
				(0.89)	(1.55)
Date FE	–	–	–	–	–
Intraday FE	✓	✓	✓	✓	✓
Observations	40799	40799	40799	40799	40799
Adj.R^2	0.324	0.328	0.307	0.307	0.331
Panel B: *Date and Intraday Fixed Effects*					
|Market Return|	0.140***	0.140***	0.140***	0.140***	0.140***
	(23.47)	(23.53)	(23.31)	(23.31)	(23.39)
Market Return^2	−0.527***	−0.527***	−0.527***	−0.526***	−0.525***
	(−2.87)	(−2.89)	(−2.85)	(−2.86)	(−2.87)
Trading Vol.	0.003***	0.003***	0.003***	0.003***	0.003***
	(11.08)	(10.95)	(10.99)	(10.91)	(10.67)
Blockchain Trans.	0.125***	0.123***	0.121***	0.120***	0.114***
	(6.34)	(6.24)	(6.17)	(6.15)	(5.89)
Search Vol.$_{\text{Level}}$	0.024***	0.048***			0.045***
	(5.58)	(7.52)			(7.14)
Search Vol.$_{\text{Dispersion}}$		0.056***			0.053***
		(5.18)			(4.98)
Reddit Posts$_{\text{Level}}$			0.057***	0.062***	0.060***
			(8.29)	(8.81)	(8.61)
Reddit Posts$_{\text{Dispersion}}$				0.068***	0.066***
				(6.86)	(6.75)
Date FE	✓	✓	✓	✓	✓
Intraday FE	✓	✓	✓	✓	✓
Observations	40798	40798	40798	40798	40798
Adj.R^2	0.549	0.550	0.550	0.551	0.552

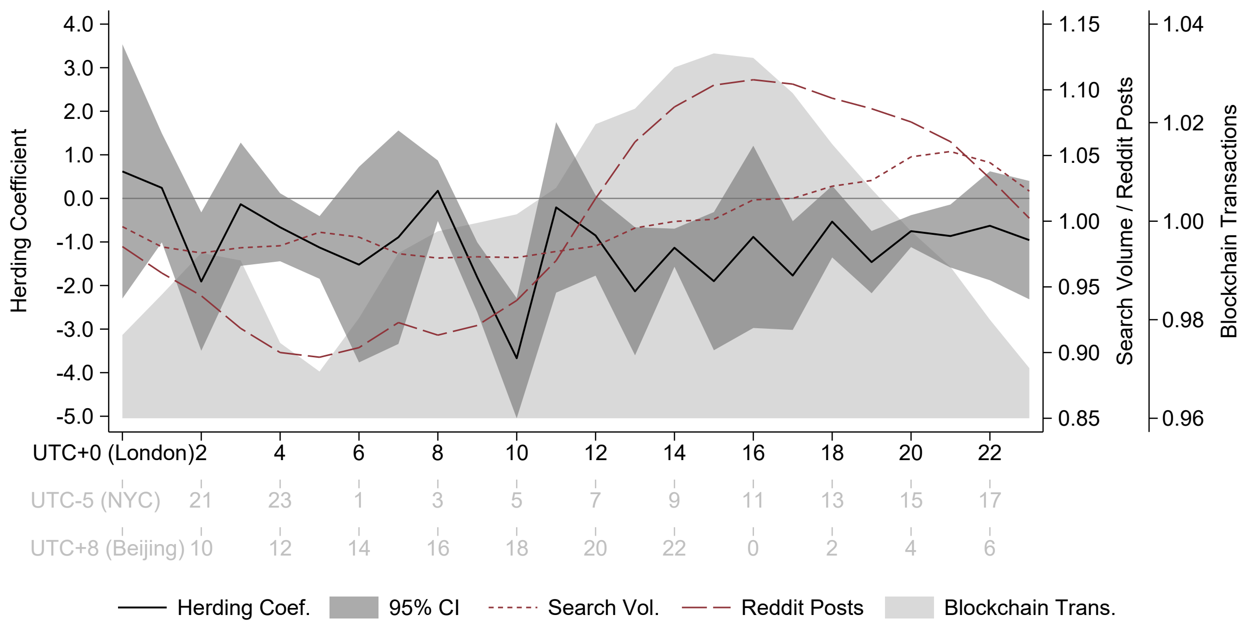
This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. In Panel A, intraday fixed effects for every one hour window of the day are included. In Panel B, date fixed effects are included in addition to the intraday fixed effects. The variables are as defined in Table 4.1, except CSAD which is here given in basis points. T-statistics with Newey and West (1987) standard errors are reported in parentheses, except for in the fixed effects model, where the standard errors are clustered by date. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Figure 4.6: Intraday Herding Patterns using other Market Indices



These graphs shows the regression coefficients $\beta_{2,h}$ from estimating Equation 4.2. In the top graph, the market index is a value weighted index of the currencies in our sample. In the bottom figure, we use Bitcoin returns as the market index. The dashed lines indicate the 95% confidence interval for the coefficient estimate based on Newey and West (1987) standard errors.

Figure 4.7: Intraday Market Return Herding Patterns



This graph shows the regression coefficients $\beta_{2,h}$ from estimating Equation 4.2. The only difference between this figure and Figure 4.2 is that control variables (i.e. *Trading Volume*, *Blockchain Transactions*, *Search* $Volume_{Level}$, *Search* $Volume_{Dispersion}$, $Reddit_{Level}$, and $Reddit_{Dispersion}$) are included during the estimation. The dark gray area indicate the 95% confidence interval for the coefficient estimate based on Newey and West (1987) standard errors. Significantly negative coefficients indicate herding. The dotted line shows the average level of search volume across the sample currencies. The long-dashed line shows the average number of posts on Reddit. Additionally, the light gray area shows the average number of transactions on the blockchains of the sample currencies. Search volume, Reddit posts, and blockchain transactions have been standardized by dividing by the respective overall means of the two time series.

Table 4.9: Baseline Herding Analysis with Extended Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Baseline Specification 1*						
|Market Return|	0.247***	0.228***	0.244***	0.197***	0.201***	0.237***
	(36.98)	(33.68)	(36.96)	(30.22)	(30.57)	(34.89)
Market Return^2	−1.308***	−1.062***	−1.241***	−0.756***	−0.823***	−1.177***
	(−6.87)	(−5.75)	(−6.63)	(−4.06)	(−4.40)	(−6.30)
Monthly Blockchain Trans.	−0.041***					
	(−4.51)					
Monthly Market Trades		0.000***				
		(7.83)				
Monthly Market Return			0.322**			
			(2.41)			
Monthly |Market Return|				0.309***		
				(16.91)		
Monthly Market Volatility					0.216***	
					(15.84)	
Monthly Market Vol.						0.001***
						(2.81)
Observations	40799	40799	40799	40799	40799	40799
Adj.R^2	0.215	0.221	0.213	0.253	0.249	0.214
Panel B: *Baseline Specification 2*						
|Market Return|	0.225***	0.223***	0.227***	0.192***	0.194***	0.226***
	(32.80)	(32.65)	(33.69)	(29.13)	(29.12)	(33.19)
Market Return^2	−1.456***	−1.152***	−1.261***	−0.797***	−0.865***	−1.348***
	(−7.55)	(−6.25)	(−6.86)	(−4.33)	(−4.71)	(−7.18)
Trading Vol.	0.004***	0.002***	0.002***	0.001***	0.001***	0.003***
	(12.09)	(4.57)	(7.62)	(3.38)	(4.49)	(9.64)
Monthly Blockchain Trans.	−0.107***					
	(−10.94)					
Monthly Market Trades		0.000***				
		(4.10)				
Monthly Market Return			0.197			
			(1.45)			
Monthly |Market Return|				0.289***		
				(15.15)		
Monthly Market Volatility					0.200***	
					(14.44)	
Monthly Market Vol.						−0.002***
						(−3.38)
Observations	40799	40799	40799	40799	40799	40799
Adj.R^2	0.236	0.224	0.223	0.255	0.252	0.224

Panel C: *Baseline Specification 3*

	(1)	(2)	(3)	(4)	(5)	(6)
|Market Return|	0.220*** (32.18)	0.222*** (32.45)	0.229*** (33.91)	0.192*** (29.42)	0.195*** (29.43)	0.227*** (33.18)
Market Return^2	−1.410*** (−7.25)	−1.205*** (−6.50)	−1.385*** (−7.35)	−0.925*** (−4.99)	−0.989*** (−5.31)	−1.374*** (−7.26)
Trading Vol.	0.004*** (11.19)	0.002*** (6.14)	0.003*** (9.75)	0.002*** (6.60)	0.003*** (7.11)	0.003*** (10.05)
Blockchain Trans.	0.234*** (8.53)	−0.100*** (−8.13)	−0.063*** (−5.95)	−0.081*** (−9.02)	−0.069*** (−7.52)	−0.060*** (−4.42)
Monthly Blockchain Trans.	−0.334*** (−12.10)					
Monthly Market Trades		0.000*** (7.46)				
Monthly Market Return			0.292** (2.08)			
Monthly |Market Return|				0.308*** (16.02)		
Monthly Market Volatility					0.206*** (14.71)	
Monthly Market Vol.						0.000 (0.13)
Observations	40799	40799	40799	40799	40799	40799
Adj.R^2	0.247	0.234	0.228	0.264	0.259	0.227

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panels A, B, and C involves the specifications of the leftmost three columns of Table 4.3, respectively. The variables are as defined in Table 4.1, except CSAD which is here given in basis points, and the extended control variables, which are introduced as follows. *Monthly Blockchain Transactions* is the monthly average of the hourly number of transactions on the blockchains of the cryptocurrencies in our sample. *Monthly Market Trades* is the monthly average of the hourly number of trades on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Volume* is the monthly average of the hourly trading volume in 1mn USD on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Return* is the monthly average of the hourly logarithmic return of the market index in percent. *Monthly |Market Return|* is the monthly average of the absolute hourly logarithmic return of the market index in percent. *Monthly Market Volatility* is the monthly standard deviation of the hourly logarithmic return of the market index in percent. T-statistics with Newey and West (1987) standard errors are reported in parentheses. The results do not include fixed effects. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.10: Herding and Investor Attention with Extended Control Variables Derived from Cryptocurrency Market

	(1)	(2)	(3)	(4)	(5)	(6)
|Market Return|	0.180*** (28.25)	0.190*** (29.74)	0.194*** (30.88)	0.188*** (29.24)	0.189*** (29.27)	0.188*** (29.62)
Market Return^2	−0.900*** (−4.79)	−0.951*** (−5.19)	−0.882*** (−4.84)	−0.867*** (−4.71)	−0.877*** (−4.76)	−1.007*** (−5.49)
Trading Vol.	0.002*** (6.19)	0.002*** (7.10)	0.001*** (4.25)	0.002*** (4.96)	0.002*** (5.02)	0.003*** (9.77)
Blockchain Trans.	0.199*** (9.36)	−0.137*** (−12.92)	−0.165*** (−16.38)	−0.147*** (−15.00)	−0.148*** (−14.66)	−0.112*** (−10.29)
Search Vol.$_{\text{Level}}$	0.193*** (25.55)	0.146*** (17.86)	0.132*** (19.27)	0.126*** (15.98)	0.128*** (16.60)	0.158*** (20.05)
Search Vol.$_{\text{Dispersion}}$	0.162*** (12.31)	0.070*** (5.56)	0.061*** (4.97)	0.055*** (4.26)	0.057*** (4.40)	0.085*** (6.70)
Reddit Posts$_{\text{Level}}$	0.010 (1.55)	0.054*** (8.44)	0.074*** (10.86)	0.057*** (8.38)	0.057*** (8.39)	0.048*** (7.63)
Reddit Posts$_{\text{Dispersion}}$	0.028* (1.85)	0.030* (1.93)	0.037** (2.45)	0.035** (2.27)	0.034** (2.21)	0.032** (2.05)
Monthly Blockchain Trans.	−0.443*** (−18.21)					
Monthly Market Trades		−0.000*** (−3.62)				
Monthly Market Return			1.094*** (8.89)			
Monthly |Market Return|				0.027 (1.26)		
Monthly Market Volatility					0.012 (0.78)	
Monthly Market Vol.						−0.005*** (−8.00)
Observations	40799	40799	40799	40799	40799	40799
Adj.R^2	0.355	0.328	0.338	0.326	0.326	0.334

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. The variables are as defined in Table 4.1, except CSAD which is here given in basis points, and the extended control variables, which are introduced as follows. *Monthly Blockchain Transactions* is the monthly average of the hourly number of transactions on the blockchains of the cryptocurrencies in our sample. *Monthly Market Trades* is the monthly average of the hourly number of trades on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Volume* is the monthly average of the hourly trading volume in 1mn USD on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Return* is the monthly average of the hourly logarithmic return of the market index in percent. *Monthly |Market Return|* is the monthly average of the absolute hourly logarithmic return of the market index in percent. *Monthly Market Volatility* is the monthly standard deviation of the hourly logarithmic return of the market index in percent. T-statistics with Newey and West (1987) standard errors are reported in parentheses. The results do not include fixed effects. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.11: Herding and Investor Attention with Extended Control Variables Derived from Stock Market

	(1)	(2)	(3)	(4)	(5)	(6)
|Market Return|	0.190*** (29.77)	0.190*** (29.75)	0.188*** (29.42)	0.189*** (29.59)	0.191*** (29.89)	0.191*** (29.74)
Market Return^2	−0.892*** (−4.86)	−0.891*** (−4.85)	−0.888*** (−4.82)	−0.887*** (−4.82)	−0.897*** (−4.90)	−0.901*** (−4.92)
Trading Vol.	0.002*** (5.08)	0.002*** (5.09)	0.002*** (5.22)	0.002*** (5.25)	0.002*** (5.08)	0.002*** (4.91)
Blockchain Trans.	−0.149*** (−15.34)	−0.149*** (−15.30)	−0.148*** (−15.18)	−0.149*** (−15.35)	−0.149*** (−15.42)	−0.139*** (−14.68)
Search Vol.$_{\text{Level}}$	0.131*** (18.85)	0.131*** (18.88)	0.133*** (18.97)	0.132*** (18.95)	0.130*** (18.72)	0.125*** (18.46)
Search Vol.$_{\text{Dispersion}}$	0.058*** (4.67)	0.058*** (4.67)	0.065*** (5.26)	0.059*** (4.78)	0.056*** (4.53)	0.055*** (4.47)
Reddit Posts$_{\text{Level}}$	0.058*** (8.73)	0.058*** (8.69)	0.061*** (9.11)	0.058*** (8.70)	0.059*** (8.81)	0.058*** (8.75)
Reddit Posts$_{\text{Dispersion}}$	0.033** (2.09)	0.033** (2.08)	0.034** (2.18)	0.032** (1.99)	0.034** (2.19)	0.033** (2.10)
Monthly Dow Jones Return	0.020 (0.32)					
Monthly S&P 500 Return		−0.003 (−0.05)				
Monthly FTSE 100 Return			−0.320*** (−4.07)			
Monthly DAX Return				−0.173*** (−3.18)		
Monthly Nikkei 225 Return					0.207*** (3.47)	
Monthly Hang Seng Return						0.424*** (7.32)
Observations	40799	40799	40799	40799	40799	40799
Adj.R^2	0.326	0.326	0.328	0.327	0.327	0.330

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Different from Table 4.10, monthly and yearly variables derived from stock market are used as control variables here. The variables are as defined in Table 4.1, except CSAD which is here given in basis points, and the extended control variables, which are introduced as follows. *Monthly Dow Jones Return* is the monthly logarithmic return of Dow Jones Industrial Average. *Monthly S&P 500 Return* is the monthly logarithmic return of S&P 500. *Monthly FTSE 100 Return* is the monthly logarithmic return of FTSE 100 Index. *Monthly DAX Return* is the monthly logarithmic return of DAX Performance Index. *Monthly Nikkei 225 Return* is the monthly logarithmic return of Nikkei 225. *Monthly Hang Seng Return* is the monthly logarithmic return of Hang Seng Index. The data source is investing.com (2024). T-statistic with Newey and West (1987) standard errors are reported in parentheses. The results do not include fixed effects. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.12: Herding in Up and Down Markets with Extended Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *Up Markets*						
|Market Return|	0.219***	0.229***	0.232***	0.228***	0.229***	0.226***
	(24.76)	(25.50)	(26.23)	(25.73)	(25.77)	(25.34)
Market Return^2	−1.541***	−1.554***	−1.487***	−1.510***	−1.518***	−1.582***
	(−4.70)	(−4.77)	(−4.60)	(−4.64)	(−4.66)	(−4.83)
Trading Vol.	0.002***	0.003***	0.002***	0.002***	0.002***	0.004***
	(4.89)	(5.43)	(3.71)	(4.19)	(4.23)	(7.20)
Blockchain Trans.	0.205***	−0.147***	−0.176***	−0.159***	−0.159***	−0.121***
	(7.81)	(−10.76)	(−13.43)	(−12.33)	(−12.12)	(−8.77)
Search Vol.$_{\text{Level}}$	0.200***	0.149***	0.136***	0.133***	0.136***	0.162***
	(20.33)	(14.34)	(15.34)	(13.41)	(13.83)	(16.13)
Search Vol.$_{\text{Dispersion}}$	0.153***	0.055***	0.047***	0.042**	0.043***	0.070***
	(9.05)	(3.44)	(2.95)	(2.57)	(2.65)	(4.34)
Reddit Posts$_{\text{Level}}$	0.005	0.052***	0.072***	0.055***	0.055***	0.045***
	(0.63)	(6.04)	(8.11)	(6.13)	(6.14)	(5.42)
Reddit Posts$_{\text{Dispersion}}$	0.031	0.037*	0.044**	0.041**	0.040*	0.038*
	(1.57)	(1.79)	(2.19)	(2.00)	(1.96)	(1.84)
Monthly Blockchain Trans.	−0.462***					
	(−15.03)					
Monthly Market Trades		−0.000***				
		(−2.77)				
Monthly Market Return			1.189***			
			(7.52)			
Monthly |Market Return|				0.008		
				(0.29)		
Monthly Market Volatility					−0.003	
					(−0.13)	
Monthly Market Vol.						−0.005***
						(−6.16)
Observations	21107	21107	21107	21107	21107	21107
Adj.R^2	0.355	0.327	0.338	0.325	0.325	0.333

	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: *Down Markets*						
|Market Return|	0.141***	0.151***	0.155***	0.147***	0.147***	0.149***
	(17.54)	(18.87)	(19.64)	(18.39)	(18.38)	(18.76)
Market Return^2	−0.174	−0.245	−0.176	−0.114	−0.126	−0.313
	(−0.83)	(−1.20)	(−0.87)	(−0.56)	(−0.62)	(−1.54)
Trading Vol.	0.001***	0.002***	0.001**	0.001***	0.001***	0.003***
	(4.34)	(5.64)	(2.52)	(3.07)	(3.12)	(8.20)
Blockchain Trans.	0.190***	−0.125***	−0.151***	−0.135***	−0.134***	−0.098***
	(7.47)	(−9.51)	(−12.19)	(−11.24)	(−10.90)	(−7.28)
Search Vol.$_{\text{Level}}$	0.186***	0.145***	0.129***	0.118***	0.120***	0.158***
	(21.30)	(14.76)	(15.85)	(12.17)	(12.72)	(16.78)
Search Vol.$_{\text{Dispersion}}$	0.168***	0.084***	0.073***	0.063***	0.065***	0.099***
	(10.06)	(5.16)	(4.60)	(3.85)	(3.99)	(6.09)
Reddit Posts$_{\text{Level}}$	0.017**	0.058***	0.076***	0.061***	0.061***	0.052***
	(2.35)	(8.03)	(9.60)	(7.81)	(7.80)	(7.27)
Reddit Posts$_{\text{Dispersion}}$	0.020	0.018	0.026	0.025	0.024	0.021
	(1.12)	(0.99)	(1.41)	(1.38)	(1.33)	(1.14)
Monthly Blockchain Trans.	−0.419***					
	(−14.69)					
Monthly Market Trades		−0.000***				
		(−3.37)				
Monthly Market Return			0.889***			
			(5.83)			
Monthly |Market Return|				0.054**		
				(2.04)		
Monthly Market Volatility					0.033*	
					(1.76)	
Monthly Market Vol.						−0.006***
						(−7.04)
Observations	19645	19645	19645	19645	19645	19645
Adj.R^2	0.371	0.345	0.351	0.343	0.343	0.353

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panels A and B only include observations during positive or negative market returns, respectively. The variables are as defined in Table 4.1, except CSAD which is here given in basis points, and the extended control variables, which are introduced as follows. *Monthly Blockchain Transactions* is the monthly average of the hourly number of transactions on the blockchains of the cryptocurrencies in our sample. *Monthly Market Trades* is the monthly average of the hourly number of trades on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Volume* is the monthly average of the hourly trading volume in 1mn USD on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Return* is the monthly average of the hourly logarithmic return of the market index in percent. *Monthly |Market Return|* is the monthly average of the absolute hourly logarithmic return of the market index in percent. *Monthly Market Volatility* is the monthly standard deviation of the hourly logarithmic return of the market index in percent. T-statistics with Newey and West (1987) standard errors are reported in parentheses. The results do not include fixed effects. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Table 4.13: Herding during High and Low Market Volatility with Extended Control Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: *High Market Volatility*						
|Market Return|	0.159***	0.171***	0.175***	0.171***	0.171***	0.168***
	(18.31)	(19.97)	(20.79)	(20.09)	(20.11)	(19.91)
Market Return^2	−0.512**	−0.587***	−0.530**	−0.531**	−0.537**	−0.619***
	(−2.19)	(−2.63)	(−2.35)	(−2.36)	(−2.38)	(−2.78)
Trading Vol.	0.002***	0.002***	0.001***	0.002***	0.002***	0.003***
	(4.63)	(5.72)	(3.32)	(3.97)	(4.00)	(7.13)
Blockchain Trans.	0.170***	−0.176***	−0.221***	−0.195***	−0.197***	−0.152***
	(5.25)	(−10.24)	(−13.58)	(−12.02)	(−11.72)	(−8.70)
Search Vol.$_{\text{Level}}$	0.218***	0.169***	0.148***	0.150***	0.152***	0.178***
	(18.51)	(13.49)	(14.13)	(11.57)	(11.90)	(14.65)
Search Vol.$_{\text{Dispersion}}$	0.167***	0.073***	0.071***	0.059***	0.060***	0.084***
	(7.99)	(3.59)	(3.64)	(2.86)	(2.90)	(4.11)
Reddit Posts$_{\text{Level}}$	−0.002	0.044***	0.075***	0.051***	0.051***	0.038***
	(−0.19)	(4.08)	(6.50)	(4.43)	(4.45)	(3.63)
Reddit Posts$_{\text{Dispersion}}$	0.036	0.041	0.066**	0.044	0.044	0.043
	(1.33)	(1.48)	(2.42)	(1.61)	(1.60)	(1.56)
Monthly Blockchain Trans.	−0.469***					
	(−12.97)					
Monthly Market Trades		−0.000***				
		(−3.66)				
Monthly Market Return			1.385***			
			(7.36)			
Monthly |Market Return|				−0.020		
				(−0.57)		
Monthly Market Volatility					−0.022	
					(−0.87)	
Monthly Market Vol.						−0.006***
						(−6.11)
Observations	16365	16365	16365	16365	16365	16365
Adj.R^2	0.343	0.319	0.332	0.316	0.316	0.325

	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: *Low Market Volatility*						
|Market Return|	0.189*** (23.29)	0.194*** (23.16)	0.197*** (24.40)	0.189*** (22.33)	0.190*** (22.43)	0.194*** (23.15)
Market Return^2	−1.685*** (−6.25)	−1.644*** (−5.97)	−1.553*** (−5.92)	−1.533*** (−5.60)	−1.557*** (−5.69)	−1.789*** (−6.43)
Trading Vol.	0.002*** (4.92)	0.002*** (4.83)	0.002*** (3.10)	0.002*** (3.70)	0.002*** (3.76)	0.003*** (7.78)
Blockchain Trans.	0.186*** (7.29)	−0.120*** (−10.34)	−0.132*** (−12.20)	−0.122*** (−11.59)	−0.121*** (−11.34)	−0.095*** (−7.91)
Search Vol.$_{\text{Level}}$	0.169*** (18.70)	0.123*** (12.54)	0.118*** (14.65)	0.108*** (12.26)	0.111*** (12.87)	0.138*** (14.69)
Search Vol.$_{\text{Dispersion}}$	0.149*** (9.80)	0.061*** (4.32)	0.054*** (3.92)	0.050*** (3.54)	0.053*** (3.75)	0.080*** (5.61)
Reddit Posts$_{\text{Level}}$	0.019*** (2.79)	0.059*** (8.74)	0.071*** (10.03)	0.059*** (8.25)	0.059*** (8.26)	0.053*** (8.02)
Reddit Posts$_{\text{Dispersion}}$	0.017 (1.17)	0.018 (1.14)	0.016 (1.07)	0.024 (1.55)	0.022 (1.47)	0.019 (1.23)
Monthly Blockchain Trans.	−0.391*** (−12.98)					
Monthly Market Trades		−0.000 (−1.11)				
Monthly Market Return			0.866*** (6.36)			
Monthly |Market Return|				0.053** (2.41)		
Monthly Market Volatility					0.028* (1.79)	
Monthly Market Vol.						−0.004*** (−5.53)
Observations	24434	24434	24434	24434	24434	24434
Adj.R^2	0.325	0.296	0.305	0.297	0.296	0.302

This table shows time-series regression results based on variations of Equation 4.1. The dependent variable is the CSAD. A significantly negative coefficient for the squared market return indicates herding. Panels A and B only include observations during high and low market return volatility, respectively. Volatility is estimated by an asymmetric power ARCH model. High market volatility periods are defined as those where the estimated volatility is larger than its moving average of the previous two weeks. The variables are as defined in Table 4.1, except CSAD which is here given in basis points, and the extended control variables, which are introduced as follows. *Monthly Blockchain Transactions* is the monthly average of the hourly number of transactions on the blockchains of the cryptocurrencies in our sample. *Monthly Market Trades* is the monthly average of the hourly number of trades on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Volume* is the monthly average of the hourly trading volume in 1mn USD on the exchange Kraken of the cryptocurrencies in our sample. *Monthly Market Return* is the monthly average of the hourly logarithmic return of the market index in percent. *Monthly |Market Return|* is the monthly average of the absolute hourly logarithmic return of the market index in percent. *Monthly Market Volatility* is the monthly standard deviation of the hourly logarithmic return of the market index in percent. T-statistics with Newey and West (1987) standard errors are reported in parentheses. The results do not include fixed effects. ***, **, * denotes significance at the 1%, 5%, 10%–level, respectively.

Bibliography

Abadi, J. and M. Brunnermeier (2018). Blockchain economics. NBER Working Paper No. 25407.

Alabi, K. (2017). Digital blockchain networks appear to be following metcalfe's law. *Electronic Commerce Research and Applications 24*, 23–29.

Alexander, C. and M. Dakos (2020). A critical investigation of cryptocurrency data and analysis. *Quantitative Finance 20*(2), 173–188.

Altxw.com (2021). Xinjiang launches coal mine safety production inspection [accessed 2021-06-02].

Andrikopoulos, P., V. Kallinterakis, M. P. Leite Ferreira, and T. Verousis (2017, oct). Intraday herding on a cross-border exchange. *International Review of Financial Analysis 53*, 25–36.

Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? The information content of Internet stock message boards. *The Journal of Finance 59*(3), 1259–1294.

Aoyagi, J. and D. Adachi (2018). Fundamental values of cryptocurrencies and blockchain technology. Working Paper.

Arnosti, N. and S. M. Weinberg (2022, jul). Bitcoin: A Natural Oligopoly. *Management Science 68*(7), 4755–4771.

Aslan, A. and A. Sensoy (2020, jul). Intraday efficiency-frequency nexus in the cryptocurrency markets. *Finance Research Letters 35*(September), 101298.

Auer, R., , and S. Claessens (2021). Cryptocurrency market reactions to regulatory news. In *The Routledge Handbook of FinTech*, pp. 455–468. Routledge.

Auer, R., J. Frost, and J. M. Vidal Pastor (2022). Miners as intermediaries: extractable value and market manipulation in crypto and DeFi. *BIS Bulletin 58*.

Ballis, A. and K. Drakos (2020, mar). Testing for herding in the cryptocurrency market. *Finance Research Letters 33*.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*(2), 785–818.

Basu, S., D. Easley, M. O'Hara, and E. Sirer (2019). Towards a functional fee market for cryptocurrencies. Working Paper.

Baumgartner, J., S. Zannettou, B. Keegan, M. Squire, and J. Blackburn (2020). The pushshift reddit dataset. *Proceedings of the 14th International AAAI Conference on Web and Social Media, ICWSM 2020*, 830–839.

Baur, D. G., D. Cahill, K. Godfrey, and Z. Liu (2019, dec). Bitcoin time-of-day, day-of-week and month-of-year effects in returns and trading volume. *Finance Research Letters 31*, 78–92.

Baur, D. G. and T. Dimpfl (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters 173*, 148–151.

Baur, D. G. and T. Dimpfl (2021). The volatility of bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics 61*, 2663—-2683.

Benetton, M., G. Compiani, and A. Morse (2022). When Cryptomining Comes to Town: High Electricity-Use Spillovers to the Local Economy. *Working Paper*.

Bernales, A., T. Verousis, and N. Voukelatos (2020). Do investors follow the herd in option markets? *Journal of Banking and Finance 119*.

Bertucci, C., L. Bertucci, J.-M. Lasry, and P.-L. Lions (2022). How Resilient is the Bitcoin Protocol? *Working Paper*.

Bhambhwani, S., S. Delikouras, and G. M. Korniotis (2019). Do fundamentals drive cryptocurrency prices? Working Paper.

Bhambhwani, S., S. Delikouras, and G. M. Korniotis (2021). Blockchain Characteristics and the Cross-Section of Cryptocurrency Returns. *Working Paper*.

Biais, B., C. Bisiere, M. Bouvard, C. Casamatta, and A. J. Menkveld (2020). Equilibrium bitcoin pricing. Working Paper.

Biais, B., C. Bisière, M. Bouvard, C. Casamatta, and A. J. Menkveld (2023, feb). Equilibrium Bitcoin Pricing. *The Journal of Finance* (0).

Bianchi, D. (2020). Cryptocurrencies as an asset class? An empirical assessment. *The Journal of Alternative Investments 23*(2), 162–179.

Bianchi, D. and M. Babiak (2021). A risk-based explanation of cryptocurrency returns. Working Paper.

Bianchi, D., M. Babiak, and A. Dickerson (2022). Trading volume and liquidity provision in cryptocurrency markets. *Journal of Banking and Finance 142*, 106547.

Binance Academy (2019, May). Orphan Block.

Binance Academy (2020a, Apr). Block header.

Binance Academy (2020b, Feb). Mempool.

Bitcoin Wiki (2018, Mar). Confirmation.

Bitcoin Wiki (2021, Feb). Block hashing algorithm.

bitFlyer EUROPE S.A. (2019, Apr). Hashrate.

Blasco, N., P. Corredor, and N. Satrústegui (2022, dec). The witching week of herding on bitcoin exchanges. *Financial Innovation 8*(1), 26.

Bohl, M. T., N. Branger, and M. Trede (2017). The case for herding is stronger than you think. *Journal of Banking and Finance 85*, 30–40.

Böhme, R., N. Christin, B. Edelman, and T. Moore (2015, may). Bitcoin: Economics, Technology, and Governance. *Journal of Economic Perspectives 29*(2), 213–238.

Bolt, W. and M. R. van Oordt (2020). On the value of virtual currencies. *Journal of Money, Credit and Banking 52*(4), 835–862.

Borri, N., D. Massacci, M. Rubin, and D. Ruzzi (2022). Crypto risk premia. Working Paper.

Bouri, E., R. Gupta, and D. Roubaud (2019, jun). Herding behaviour in cryptocurrencies. *Finance Research Letters 29*, 216–221.

Bouri, E., C. K. M. Lau, B. Lucey, and D. Roubaud (2019). Trading volume and the predictability of return and volatility in the cryptocurrency market. *Finance Research Letters 29*, 340–346.

Brauneis, A., R. Mestel, R. Riordan, and E. Theissen (2021). How to measure the liquidity of cryptocurrency markets? *Journal of Banking and Finance 124*, 106041.

Brauneis, A., R. Mestel, and E. Theissen (2023). The Crypto World Trades at Tea Time. Intraday Evidence from Centralized Exchanges across the Globe. *Working Paper*.

Browne, R. and M. Sigalos (2024, Jan). United states acts as top cop - setting the crypto standards for the world.

Bystroem, H. and D. Krygier (2018). What drives bitcoin volatility. *Working Paper*.

Cai, C. X. and R. Zhao (2021). Salience theory and cryptocurrency returns. Working Paper.

Cai, F., S. Han, D. Li, and Y. Li (2019, oct). Institutional herding and its price impact: Evidence from the corporate bond market. *Journal of Financial Economics 131*(1), 139–167.

Cambridge Centre for Alternative Finance (2021). Cambridge Bitcoin Electricity Consumption Index [accessed 2021-06-02].

Capponi, A., S. Olafsson, and H. Alsabah (2021). Proof-of-Work Cryptocurrencies: Does Mining Technology Undermine Decentralization? *Management Science*.

Carlin, B. I., F. A. Longstaff, and K. Matoba (2014). Disagreement and asset prices. *Journal of Financial Economics 114*(2), 226–238.

Catania, L. and S. Grassi (2022). Forecasting cryptocurrency volatility. *International Journal of Forecasting 38*(3), 878–894.

Chang, E. C., J. W. Cheng, and A. Khorana (2000, oct). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking and Finance 24*(10), 1651–1679.

Chen, C. and L. Liu (2022, may). How effective is China's cryptocurrency trading ban? *Finance Research Letters 46*, 102429.

Chen, T., R. K. Larson, and H. Mo (2022, jul). Investor Herding and Price Informativeness in Global Markets: Evidence from Earnings Announcements. *Journal of Behavioral Finance*, 1–19.

Chiang, T. C. and D. Zheng (2010). An empirical analysis of herd behavior in global stock markets. *Journal of Banking and Finance 34*(8), 1911–1921.

Choi, K.-H., S. H. Kang, and S.-M. Yoon (2022, may). Herding behaviour in Korea's cryptocurrency market. *Applied Economics 54*(24), 2795–2809.

Choi, N. and R. W. Sias (2009). Institutional industry herding. *Journal of Financial Economics 94*(3), 469–491.

Christie, W. G. and R. D. Huang (1995, jul). Following the Pied Piper: Do Individual Returns Herd around the Market? *Financial Analysts Journal 51*(4), 31–37.

CNBC (2021). Iran bans bitcoin mining as its cities suffer blackouts and power shortages [accessed 2021-07-27].

CNBC (2022). Kazakhstan's deadly protests hit bitcoin, as the world's second-biggest mining hub shuts down [accessed 2022-08-16].

Coinbase (2020, Dec). What is mining?

Coinbase (2024, Mar). What is Bitcoin halving?

Cong, L. W., Z. He, and J. Li (2021). Decentralized Mining in Centralized Pools. *The Review of Financial Studies 34*(3), 1191–1235.

Cong, L. W., G. A. Karolyi, K. Tang, and W. Zhao (2022). Value premium, network adoption, and factor pricing of crypto assets. Working Paper.

Conrad, C., A. Custovic, and E. Ghysels (2018). Long-and short-term cryptocurrency volatility components: A garch-midas analysis. *Journal of Risk and Financial Management 11*(2), 23.

Corbet, S., C. Larkin, and B. Lucey (2020). The contagion effects of the COVID-19 pandemic: Evidence from gold and cryptocurrencies. *Finance Research Letters 35*, 101554.

Corbet, S., B. Lucey, and L. Yarovaya (2021, mar). Bitcoin-energy markets interrelationships - New evidence. *Resources Policy 70*, 101916.

Corwin, S. A. and P. Schultz (2012, apr). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance 67*(2), 719–760.

Coskun, E. A., C. K. M. Lau, and H. Kahyaoglu (2020). Uncertainty and herding behavior: evidence from cryptocurrencies. *Research in International Business and Finance 54*(2), 101284.

Cousins, K., H. Subramanian, and P. Esmaeilzadeh (2019). A value-sensitive design perspective of cryptocurrencies: A research agenda. *Communications of the Association for Information Systems 45*(27), 511–547.

Da, Z., J. Engelberg, and P. Gao (2011, oct). In Search of Attention. *The Journal of Finance 66*(5), 1461–1499.

da Gama Silva, P. V. J., M. C. Klotzle, A. C. F. Pinto, and L. L. Gomes (2019, jun). Herding behavior and contagion in the cryptocurrency market. *Journal of Behavioral and Experimental Finance 22*, 41–50.

de Vries, A. (2020). Bitcoin's energy consumption is underestimated: A market dynamics approach. *Energy Research and Social Science 70*(July), 101721.

de Vries, A., U. Gallersdörfer, L. Klaaßen, and C. Stoll (2022, mar). Revisiting Bitcoin's carbon footprint. *Joule 6*(3), 498–502.

Delgado-Mohatar, O., M. Felis-Rota, and C. Fernández-Herraiz (2019). The Bitcoin mining breakdown: Is mining still profitable? *Economics Letters 184*, 108492.

Digiconomist.net (2021). Amidst avid greenwashing, Xinjiang blackout exposes the horrid reality of Bitcoin mining [accessed 2021-06-02].

Ding, Z., C. W. J. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Dittmar, L. and A. Praktiknjo (2019). Could Bitcoin emissions push global warming above 2 °C? *Nature Climate Change 9*(9), 656–657.

Drake, M. S., J. Jennings, D. T. Roulstone, and J. R. Thornock (2017, sep). The Comovement of Investor Attention. *Management Science 63*(9), 2847–2867.

Dwyer, G. P. (2015). The economics of bitcoin and similar private digital currencies. *Journal of Financial Stability 17*, 81–91.

Dyhrberg, A. H., S. Foley, and J. Svec (2018, oct). How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets. *Economics Letters 171*, 140–143.

D'Amato, V., S. Levantesi, and G. Piscopo (2022). Deep learning in predicting cryptocurrency volatility. *Physica A: Statistical Mechanics and its Applications 596*, 127158.

Easley, D., M. O'Hara, and S. Basu (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics 134*(1), 91–109.

Easley, D., M. O'Hara, and S. Basu (2019). From mining to markets: The evolution of bitcoin transaction fees. *Journal of Financial Economics 134*(1), 91–109.

Eross, A., F. McGroarty, A. Urquhart, and S. Wolfe (2019, oct). The intraday dynamics of bitcoin. *Research in International Business and Finance 49*(June 2018), 71–81.

ESMA (2023). Markets in crypto-assets regulation (mica).

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Ferreira, D., J. Li, and R. Nikolowa (2022, aug). Corporate Capture of Blockchain Governance. *The Review of Financial Studies*.

Finck, M. (2018). Blockchains: Regulating the unknown. *German Law Journal 19*(4), 665–692.

Fink, C. and T. Johann (2014). May i have your attention, please: The market microstructure of investor attention. *Please: The Market Microstructure of Investor Attention (September 17, 2014)*.

Fortune.com (2021). How much Bitcoin comes from dirty coal? A flooded mine in China just spotlighted the issue [accessed 2021-06-02].

Gao, C. and B. Mizrach (2016). Market quality breakdowns in equities. *Journal of Financial Markets 28*, 1–23.

Garriga, M., S. Dalla Palma, M. Arias, A. Renzis, R. Pareschi, and D. A. Tamburri (2020). Blockchain and cryptocurrencies: A classification and comparison of architecture drivers. *Concurrency and Computation: Practice and Experience 33*(8), e5992.

Gebka, B. and M. E. Wohar (2013). International herding: Does it differ across sectors? *Journal of International Financial Markets, Institutions and Money 23*(1), 55–84.

Gleason, K. C., I. Mathur, and M. A. Peterson (2004, dec). Analysis of intraday herding behavior among the sector ETFs. *Journal of Empirical Finance 11*(5), 681–694.

Gurdgiev, C. and D. O'Loughlin (2020, mar). Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty. *Journal of Behavioral and Experimental Finance 25*.

Hafner, C. M. (2020). Testing for bubbles in cryptocurrencies with time-varying volatility. *Journal of Financial Econometrics 18*(2), 233—-249.

Hajdarbegovic, N. (2014, Jan). Bitcoin miners ditch Ghash.io pool over fears of 51% attack.

Härdle, W. K., C. R. Harvey, and R. C. G. Reule (2020, mar). Understanding Cryptocurrencies. *Journal of Financial Econometrics 18*(2), 181–208.

Hautsch, N., C. Scheuch, and S. Voigt (2021). Building Trust Takes Time: Limits to Arbitrage in Blockchain-Based Markets. *Working Paper*.

Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics 34*(7), 1308–1321.

Hirshleifer, D. (2015, dec). Behavioral Finance. *Annual Review of Financial Economics 7*(1), 133–159.

Hirshleifer, D. and S. H. Teoh (2003, dec). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1-3), 337–386.

Hoang, L. T. and D. G. Baur (2021, jul). How stable are stablecoins? *The European Journal of Finance*, 1–17.

Hoang, L. T. and D. G. Baur (2022, nov). Loaded for bear: Bitcoin private wallets, exchange reserves and prices. *Journal of Banking and Finance 144*, 106622.

Hoenicke, J. (2021). Johoe's Bitcoin Mempool Statistics.

Hsieh, S. F. (2013). Individual and institutional herding and the impact on stock returns: Evidence from Taiwan stock market. *International Review of Financial Analysis 29*, 175–188.

Hsieh, S. F., C. Y. Chan, and M. C. Wang (2020, dec). Retail investor attention and herding behavior. *Journal of Empirical Finance 59*, 109–132.

Hu, A. S., C. A. Parlour, and U. Rajan (2019). Cryptocurrencies: Stylized facts on a new investible instrument. *Financial Management 48*(4), 1049–1068.

Hu, B., T. McInish, J. Miller, and L. Zeng (2019, mar). Intraday price behavior of cryptocurrencies. *Finance Research Letters 28*(June), 337–342.

Huberman, G., J. D. Leshno, and C. C. Moallemi (2021). Monopoly without a monopolist: An economic analysis of the bitcoin payment system. *The Review of Economic Studies 88*(6), 3011–3040.

investing.com (2024). Investing.com - Stock Market Quotes & Fiancial News.

Irresberger, F., K. John, P. Mueller, and F. Saleh (2020). The public blockchain ecosystem: An empirical analysis. NYU Stern School of Business Working Paper.

Irresberger, F., K. John, and F. Saleh (2021). The Public Blockchain Ecosystem: An Empirical Analysis. *Working Paper* (212).

Jafarinejad, M. and H. Sakaki (2018, mar). The Role of Investor Attention in Bitcoin Mania. *Working Paper*.

Ji, Q., E. Bouri, D. Roubaud, and L. Kristoufek (2019, jun). Information interdependence among energy, cryptocurrency and major commodity markets. *Energy Economics 81*(December 2013), 1042–1055.

Jo, H., H. Park, and H. Shefrin (2020). Bitcoin and sentiment. *Journal of Futures Markets 40*(12), 1861–1879.

John, K., T. Rivera, and F. Saleh (2021). Economic Implications of Scaling Blockchains: Why the Consensus Protocol Matters. *Working Paper*.

Joseph, K., M. Babajide Wintoki, and Z. Zhang (2011, oct). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting 27*(4), 1116–1127.

Kaiser, L. and S. Stöckl (2020, mar). Cryptocurrencies: Herding and the transfer currency. *Finance Research Letters 33*, 101214.

Kallinterakis, V. and Y. Wang (2019, dec). Do investors herd in cryptocurrencies – and why? *Research in International Business and Finance 50*, 240–245.

Karmakar, S., R. Demirer, and R. Gupta (2021, oct). Bitcoin mining activity and volatility dynamics in the power market. *Economics Letters*, 110111.

Karnaukh, N., A. Ranaldo, and P. Söderlind (2015). Understanding FX liquidity. *The Review of Financial Studies 28*(11), 3073–3108.

Katsiampa, P. (2019). An empirical investigation of volatility dynamics in the cryptocurrency market. *Research in International Business and Finance 50*, 322–335.

Katsiampa, P., S. Corbet, and B. Lucey (2019). Volatility spillover effects in leading cryptocurrencies: A bekk-mgarch analysis. *Finance Research Letters 29*, 68–74.

Kim, S. T. (2020). Bitcoin dilemma: Is popularity destroying value? *Finance Research Letters 33*(July).

Kim, T. (2015). The predecessors of bitcoin and their implications for the prospect of virtual currencies. *PLOS ONE 10*(4), e0123071.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun (2017). The Flash Crash: High-Frequency Trading in an Electronic Market. *Journal of Finance 72*(3), 967–998.

Koenraadt, J. and E. Leung (2022). Investor reactions to crypto token regulation. *European Accounting Review*.

Kremer, S. and D. Nautz (2013). Causes and consequences of short-term institutional herding. *Journal of Banking and Finance 37*(5), 1676–1686.

Kristoufek, L. and M. Vosvrda (2019). Cryptocurrencies market efficiency ranking: Not so straightforward. *Physica A: Statistical Mechanics and its Applications 531*, 120853.

Kyle, A. S. and A. A. Obizhaeva (2016). Market Microstructure Invariance: Empirical Hypotheses. *Econometrica 84*(4), 1345–1404.

Köchling, G., J. Müller, and P. N. Posch (2019). Does the introduction of futures improve the efficiency of bitcoin? *Finance Research Letters 30*, 367–370.

Lehar, A. and C. A. Parlour (2022). Miner Collusion and the Bitcoin Protocol. *Working Paper*.

Leong, M. and S. Kwok (2023). The pricing of jump and diffusive risks in the cross-section of cryptocurrency returns. *Journal of Empirical Finance 74*, 101420.

Li, L. and P. Miu (2023). Are cryptocurrencies a safe haven for stock investors? a regime-switching approach. *Journal of Empirical Finance 70*, 367–385.

Liebi, L. J. (2022). Is there a value premium in cryptoasset markets? *Economic Modelling 109*, 105777.

Liu, W., X. Liang, and G. Cui (2020). Common risk factors in the returns on cryptocurrencies. *Economic Modelling 86*, 299–305.

Liu, Y. and A. Tsyvinski (2021). Risks and returns of cryptocurrency. *The Review of Financial Studies 34*(6), 2689–2727.

Liu, Y., A. Tsyvinski, and X. Wu (2022). Common risk factors in cryptocurrency. *The Journal of Finance 77*(2), 1133–1177.

Liu, Z., S. Lu, B. Li, and S. Wang (2023, jun). Time series momentum and reversal: Intraday information from realized semivariance. *Journal of Empirical Finance 72*, 54–77.

Makarov, I. and A. Schoar (2020). Trading and arbitrage in cryptocurrency markets. *Journal of Financial Economics 135*(2), 293–319.

Makarov, I. and A. Schoar (2022). Blockchain Analysis of the Bitcoin Market. *Working Paper*.

Mandaci, P. E. and E. C. Cagli (2022, may). Herding intensity and volatility in cryptocurrency markets during the COVID-19. *Finance Research Letters 46*(February), 102382.

Menkveld, A. J. and B. Z. Yueshen (2019). The flash crash: A cautionary tale about highly fragmented markets. *Management Science 65*(10), 4470–4488.

Meshcheryakov, A. and D. B. Winters (2020, nov). Retail investor attention and the limit order book: Intraday analysis of attention-based trading. *International Review of Financial Analysis*, 101627.

Metcalfe, B. (2013). Metcalfe's law after 40 years of ethernet. *Computer 46*(12), 26–31.

Mohamad, A. and S. Stavroyiannis (2022, sep). Do birds of a feather flock together? Evidence from time-varying herding behaviour of bitcoin and foreign exchange majors during Covid-19. *Journal of International Financial Markets, Institutions and Money 80*, 101646.

Mora, C., R. L. Rollins, K. Taladay, M. B. Kantar, M. K. Chock, M. Shimada, and E. C. Franklin (2018). Bitcoin emissions alone could push global warming above 2°C. *Nature Climate Change 8*(11), 931–933.

Mueller-Bloch, C., J. V. Andersen, J. Spasovski, and J. Hahn (2022, sep). Understanding decentralization of decision-making power in proof-of-stake blockchains: an agent-based simulation approach. *European Journal of Information Systems*, 1–20.

Nabilou, H. and A. Prüm (2019). Ignorance, debt and cryptocurrencies: The old and the new in the law and economics of concurrent currencies. *Journal of Financial Regulation 5*(1), 29–63.

Nakamoto, S. (2008a). Bitcoin: A peer-to-peer electronic cash system. Online, available from https://bitcoin.org/bitcoin.pdf.

Nakamoto, S. (2008b). Bitcoin: A peer-to-peer electronic cash system.

Neumüller, A. (2023, Dec). Ethereum's climate impact: A contemporary and historical perspective.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Noda, A. (2021). On the evolution of cryptocurrency market efficiency. *Applied Economics Letters 28*(6), 433–439.

Pagnotta, E. S. (2022). Decentralizing money: Bitcoin prices and blockchain security. *The Review of Financial Studies 35*(2), 866–907.

Pagnotta, E. S. and A. Buraschi (2018). An equilibrium valuation of bitcoin and decentralized network assets. Working Paper.

Papadamou, S., N. A. Kyriazis, P. Tzeremes, and S. Corbet (2021, jun). Herding behaviour and price convergence clubs in cryptocurrencies during bull and bear markets. *Journal of Behavioral and Experimental Finance 30*, 100469.

Park, B. J. (2011, oct). Asymmetric herding as a source of asymmetric return volatility. *Journal of Banking and Finance 35*(10), 2657–2665.

Pástor, Ľ., R. F. Stambaugh, and L. A. Taylor (2015). Scale and skill in active management. *Journal of Financial Economics 116*(1), 23–45.

Petukhina, A. A., R. C. Reule, and W. K. Härdle (2021). Rise of the machines? Intraday high-frequency trading patterns of cryptocurrencies. *European Journal of Finance 27*(1-2), 8–30.

Pfautsch, F., N. Schubert, C. Orglmeister, M. Gebhart, P. Habermann, and B. Juurlink (2020). The evolution of secure hash algorithms. *PARS-Mitteilungen 35*(1), 5–15.

Pham, S. D., T. T. T. Nguyen, and H. X. Do (2022, aug). Dynamic volatility connectedness between thermal coal futures and major cryptocurrencies: Evidence from China. *Energy Economics 112*, 106114.

Philippas, D., N. Philippas, P. Tziogkidis, and H. Rjiba (2020, mar). Signal-herding in cryptocurrencies. *Journal of International Financial Markets, Institutions and Money 65*, 101191.

Phillips, R. C. and D. Gorse (2018). Cryptocurrency price drivers: Wavelet coherence analysis revisited. *PLoS ONE 13*(4), 1–21.

Piccoli, P. and M. Chaudhury (2019). Dynamic Relationship of Cryptocurrency Prices and Investor Attention. *Working Paper*.

Polasik, M., A. I. Piotrowska, T. P. Wisniewski, R. Kotkowski, and G. Lightfoot (2015). Price fluctuations and the use of bitcoin: An empirical inquiry. *International Journal of Electronic Commerce 20*(1), 9–49.

Raimundo Júnior, G. d. S., R. B. Palazzi, R. d. S. Tavares, and M. C. Klotzle (2022, jan). Market Stress and Herding: A New Approach to the Cryptocurrency Market. *Journal of Behavioral Finance 23*(1), 43–57.

Ren, B. and B. Lucey (2022). A clean, green haven?—Examining the relationship between clean energy, clean and dirty cryptocurrencies. *Energy Economics 109*(March), 105951.

Roll, R., E. Schwartz, and A. Subrahmanyam (2007). Liquidity and the law of one price: The case of the futures-cash basis. *The Journal of Finance 62*(5), 2201–2234.

Roşu, I. and F. Saleh (2021, feb). Evolution of Shares in a Proof-of-Stake Cryptocurrency. *Management Science 67*(2), 661–672.

Sabherwal, S., S. K. Sarkar, and Y. Zhang (2011, nov). Do Internet Stock Message Boards Influence Trading? Evidence from Heavily Discussed Stocks with No Fundamental News. *Journal of Business Finance and Accounting 38*(9-10), 1209–1237.

Saleh, F. (2018). Volatility and welfare in a crypto economy. *Working Paper*.

Saleh, F. (2021). Blockchain without waste: Proof-of-stake. *The Review of Financial Studies 34*(3), 1156–1190.

Scharnowski, S. (2021a, jan). Understanding Bitcoin liquidity. *Finance Research Letters 38*, 101477.

Scharnowski, S. (2021b, jan). Understanding Bitcoin liquidity. *Finance Research Letters 38*, 101477.

Scharnowski, S. and Y. Shi (2024). Intraday herding and attention around the clock. *Journal of Behavioral and Experimental Finance*, 100894.

Schilling, L. and H. Uhlig (2019). Some simple bitcoin economics. *Journal of Monetary Economics 106*, 16–26.

See-To, E. W. and Y. Yang (2017). Market sentiment dispersion and its effects on stock return and volatility. *Electronic Markets 27*(3), 283–296.

Shams, A. (2020). The structure of cryptocurrency returns. Fisher College of Business Working Paper No. 2020-03-011.

Sharif, A., M. Brahim, E. Dogan, and P. Tzeremes (2023, apr). Analysis of the spillover effects between green economy, clean and dirty cryptocurrencies. *Energy Economics 120*, 106594.

Shleifer, A. and R. W. Vishny (1997, mar). The Limits of Arbitrage. *The Journal of Finance 52*(1), 35–55.

Sias, R. W. (2004). Institutional Herding. *Review of Financial Studies 17*(1), 165–206.

Sibande, X., R. Gupta, R. Demirer, and E. Bouri (2023, jan). Investor Sentiment and (Anti) Herding in the Currency Market: Evidence from Twitter Feed Data. *Journal of Behavioral Finance 24*(1), 56–72.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics 69*(1), 99.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review 63*(2), 129–138.

Sockin, M. and W. Xiong (2023). A model of cryptocurrencies. *Management Science 69*(11), 6684–6707.

Sokolov, K. (2021, aug). Ransomware activity and blockchain congestion. *Journal of Financial Economics 141*(2), 771–782.

Stavroyiannis, S., V. Babalos, S. Bekiros, and S. Lahmiri (2019). Is anti-herding behavior spurious? *Finance Research Letters 29*, 379–383.

Stinner, J. and M. Tyrell (2022). Proof-Of-Work Consensus under Exogenous Distress: Evidence from Mining Shocks in the Bitcoin Ecosystem. *Working Paper*.

Theblockcrypto.com (2021). Bitcoin's hash rate set to regain loss as Xinjiang miners gradually resume operations [accessed 2021-06-02].

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267—-288.

Trimborn, S. and W. K. Härdle (2018). CRIX - An index for cryptocurrencies. *Journal of Empirical Finance 49*, 107–122.

Tversky, A. and D. Kahneman (1974, sep). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–1131.

Urquhart, A. (2017). The volatility of bitcoin. *Working Paper*.

Vidal-Tomás, D., A. M. Ibáñez, and J. E. Farinós (2019, sep). Herding in the cryptocurrency market: CSSD and CSAD approaches. *Finance Research Letters 30*, 181–186.

Wang, S. and J.-P. Vergne (2017). Buzz factor or innovation potential: What explains cryptocurrencies' returns? *PLOS ONE 12*(1), e0169556.

Wang, X., M. H. Kim, and S. Suardi (2022, aug). Herding and China's market-wide circuit breaker. *Journal of Banking and Finance 141*, 106533.

Wei, W. C. (2018). Liquidity and market efficiency in cryptocurrencies. *Economics Letters 168*(1), 21–24.

Yarovaya, L., R. Matkovskyy, and A. Jalan (2021, mar). The effects of a "black swan" event (COVID-19) on herding behavior in cryptocurrency markets. *Journal of International Financial Markets, Institutions and Money*, 101321.

Yen, K.-C. and H.-P. Cheng (2021). Economic policy uncertainty and cryptocurrency volatility. *Finance Research Letters 38*, 101428.

Yi, S., Z. Xu, and G.-J. Wang (2018). Volatility connectedness in the cryptocurrency market: Is bitcoin a dominant cryptocurrency? *International Review of Financial AnalysiS 60*, 98–114.

Yousaf, I. and L. Yarovaya (2022, dec). Herding behavior in conventional cryptocurrency market, non-fungible tokens, and DeFi assets. *Finance Research Letters 50*, 103299.

Youssef, M. (2022a, jan). Do Oil Prices and Financial Indicators Drive the Herding Behavior in Commodity Markets? *Journal of Behavioral Finance 23*(1), 58–72.

Youssef, M. (2022b). What Drives Herding Behavior in the Cryptocurrency Market? *Journal of Behavioral Finance 23*(2), 230–239.

Youssef, M. and S. S. Waked (2022, nov). Herding behavior in the cryptocurrency market during COVID-19 pandemic: The role of media coverage. *The North American Journal of Economics and Finance 62*, 101752.

Zhang, W., Y. Li, X. Xiong, and P. Wang (2021). Downside risk and the cross-section of cryptocurrency returns. *Journal of Banking and Finance 133*, 106246.

Zhang, W. and P. Wang (2020, jul). Investor attention and the pricing of cryptocurrency market. *Evolutionary and Institutional Economics Review 17*(2), 445–468.

Zheng, Z., K. Tang, Y. Liu, and M. Guo (2021). Gender and Herding. *Journal of Empirical Finance.*

Zimmerman, P. (2020a). Blockchain structure and cryptocurrency prices. Bank of England Working Paper No. 855.

Zimmerman, P. (2020b). Blockchain structure and cryptocurrency prices. *Bank of England Staff Working Paper No . 855* (855), 1–71.

Curriculum Vitae

September, 2019 - January, 2025	**University of Mannheim** Mannheim, Germany *PhD - Candidate in Finance*
September, 2017 - July, 2019	**University of Mannheim** Mannheim, Germany *Mannheim Master in Business Research in Marketing (Master of Science)*
September, 2014 - June, 2017	**University of Mannheim** Mannheim, Germany *Bachelor of Science in Business informatics*
July, 2014	**International School Stuttgart** Stuttgart, Germany *IB-Diploma (A-Levels equivalent)*