CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Liquidity and Predictability of Cryptoassets

Master's Thesis

Author: Bc. Viktória Mjartanová
Study program: Economics and Finance
Supervisor: prof. PhDr. Ladislav Krištoufek, Ph.D.
Year of defense: 2021

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, September 25, 2020

Bc. Viktória Mjartanová

Abstract

The relationship between liquidity and return predictability may be an important aspect to consider when investing in cryptoassets. We examine this relation using both cross-sectional as well as panel data. First, we calculate a set of predictability measures and aggregate the results into four variables. We then regress the predictability variables on a set of controls and two measures of liquidity, specifically the Amihud illiquidity ratio and the Corwin-Schultz spread estimate. The other independent variables include the logarithm of volume, turnover ratio and Garman-Klass volatility. Results from the cross-sectional analysis indicate that liquidity negatively impacts the degree of return predictability. Moreover, findings from a subset of panel data, including only 50 cryptoassets with the largest market capitalization, provide some evidence in favor of this relationship. Results from full panel data, however, present contradictory evidence. For these regressions, liquidity is found to be either insignificant or to possess a positive impact on the degree of return predictability. Altogether, we obtain mixed evidence about the effect of cryptoasset liquidity on return predictability.

JEL Classification	C53, C58, G14
Keywords	Cryptoassets, Predictability, Liquidity, Panel data

Title	Liquidity and Predictability of Cryptoassets

Author's e-mail	`viktoria.mjartan@gmail.com`
Supervisor's e-mail	`ladislav.kristoufek@fsv.cuni.cz`

Abstrakt

Pri investovaní do kryptoaktív môže byť dôležitým aspektom vplyv likvidity na prediktabilitu výnosu. Regresnou analýzou na prierezových a panelových dátach som skúmala dopad likvidity na prediktabilitu výnosu kryptoaktív. Výsledky vypočítaných testov na prediktabilitu som agregovala do štyroch závislých premenných. Nezávislé premenné zahŕňajú Garman-Klass volatilitu, pomer obratu, logaritmus objemu a dve miery likvidity (Amihud a Corwin-Schultz). Výsledky prierezovej analýzy poukazujú na negatívny dopad likvidity na prediktabilitu výnosu. Tento dopad čiastočne podporujú výsledky panelovej analýzy na podskupine dát, v ktorej sa nachádzalo 50 kryptoaktív s najväčšou kapitalizáciou. Avšak výsledky panelovej analýzy na kompletnej sade dát poskytujú protichodné zistenia. V týchto regresiách je likvidita buď nesignifikantná, alebo má pozitívny vplyv na prediktabilitu výnosu. V súhrne, výsledky regresnej analýzy poukazujú na rozličné zistenia o dopade likvidity na prediktabilitu výnosu kryptoaktív.

Klasifikácia JEL	C53, C58, G14
Kľúčové slová	Kryptoaktíva, Prediktabilita, Likvidita, Panelové dáta

Názov práce	Likvidita a prediktabilita kryptoaktív
E-mail autora	viktoria.mjartan@gmail.com
E-mail vedúceho práce	ladislav.kristoufek@fsv.cuni.cz

Acknowledgements

Prof. Krištoufek greatly supported me in writing and completing this thesis. I would like to sincerely thank him for his help, advice and encouragement.

Bibliographic Record

Mjartanová, Viktória: *Liquidity and Predictability of Cryptoassets.* Master's Thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2020, pages 73. Advisor: prof. PhDr. Ladislav Krištoufek, Ph.D.

Contents

List of Tables **viii**

Acronyms **ix**

Thesis Proposal **x**

1 Introduction **1**

2 Cryptoassets **4**

3 Literature Review **9**

3.1 Predictability, Liquidity and Market Efficiency 9

3.2 Predictability and Liquidity of Cryptoassets 11

3.3 Empirical Approaches 13

4 Methodology **19**

4.1 Predictability Measures 19

4.2 Liquidity Measures 23

4.3 Econometric Approach 26

5 Data **31**

5.1 Data Collection 31

5.2 Summary Statistics 32

6 Econometric Results **35**

6.1 Results for Cross-Section Models 35

6.2 Results for Panel Data 38

6.3 Further Discussion and Limitations 45

7 Conclusion **49**

References **52**

Appendix **I**

List of Tables

3.1 Measures Used in Cryptoasset Literature 14

5.1 Summary Statistics for Predictability Measures 32
5.2 Summary Statistics 33

6.1 Baseline Cross-Section Models 36
6.2 Cross-Section Models 37
6.3 Baseline Panel Data Models 39
6.4 Panel Data Models 41
6.5 Baseline Panel Data Models (Subset) 43
6.6 Panel Data Models (Subset) 44

A1 List of Cryptoassets I
A2 Average Partial Effects for Logit Models II
A3 Cross-Section Models with Quadratic Terms III
A4 Panel Data Models with Quadratic Terms IV
A5 Panel Data Models with Quadratic Terms (Subset) V

Acronyms

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

DLT Distributed Ledger Technology

GPH Geweke and Porter-Hudak

OLS Ordinary Least Squares

VIF Variance Inflation Factor

Master's Thesis Proposal

Author	Bc. Viktória Mjartanová
Supervisor	prof. PhDr. Ladislav Krištoufek, Ph.D.
Proposed topic	Liquidity and Predictability of Cryptoassets

Motivation Cryptoassets are an exciting and interesting asset class for potential investment. The European Central Bank Cryptoassets Task Force (2019) reports that there exist about 1900 cryptoassets. This amount may be especially striking, when considering that only 7 cryptoassets were reported to exist in April 2013 (European Central Bank Cryptoassets Task Force, 2019). Crypto-markets are still growing and continue to interest investors. It may thus be important for them to devise potentially profit-generating trading strategies. By analyzing the relationship between cryptoasset liquidity and return predictability, we may provide insights about the crypto-markets. These findings may then be used to better understand crypto-markets and thus create more befitting trading strategies.

Furthermore, although numerous studies analyze cryptoasset predictability, only several also account for its relationship with liquidity. For example, Brauneis and Mestel (2018) as well as Wei (2018) find evidence in support of a negative relationship between cryptoasset liquidity and return predictability. Nevertheless, considering that to the best of our knowledge, we are not aware of a study that would examine this relation using panel data regressions, there remains ample room for more research.

Hypotheses

Hypothesis #1: Liquidity has a negative effect on the degree of return predictability for cross-sectional data.

Hypothesis #2: Liquidity has a negative effect on the degree of return predictability for the full panel data set.

Hypothesis #3: Liquidity has a negative effect on the degree of return predictability for the panel data subset.

Methodology Cryptoasset market data is obtained from CoinMarketCap. Following previous literature, namely Brauneis and Mestel (2018), we calculate various predictability measures. We specifically include the Bartels test, runs test, difference sign test, Ljung-Box test, turning point test, two Lo-MacKinlay variance ratio tests for different holding periods and the Geweke and Porter-Hudak (GPH) estimate. Four dependent variables are created based on the p-values from the seven tests and the absolute value of the GPH estimate. The variables are average cryptoasset rank, average p-value, the total number of tests that cannot reject the null hypothesis and a binary variable, equal to 1, when a cryptoasset has at least four tests for which the null hypothesis cannot be rejected.

Liquidity may be measured by the Amihud illiquidity ratio and the Corwin-Schultz spread estimate. Following the regression specification by Brauneis and Mestel (2018), other independent variables include the Garman-Klass volatility, logarithm of volume, logarithm of market capitalization and the turnover ratio. To quantify the relationship between liquidity and predictability we regress the dependent variables on the independent variables using ordinary least squares and the logit for cross-sectional data as well as the fixed effects linear and logistic models on the full and subset panel data sets. We, furthermore, check for potential quadratic effects.

Expected Contribution To the best of our knowledge only a few studies investigate the relationship between cryptoasset liquidity and return predictability. Two notable examples are Brauneis and Mestel (2018) as well as Wei (2018). Nevertheless, at the time of writing, no existing study examines this relation using panel data regressions. In our analysis, we first build on the study by Brauneis and Mestel (2018) and carry out cross-sectional regressions using a different time period as well as more cryptoassets. Moreover, we also include three new dependent variables. Furthermore, we perform analogous regressions on panel data. Specifically, we calculate the linear fixed effects and the logistic fixed effects models. Our results may then be used to improve the understanding of crypto-markets and thus potentially help devise trading strategies.

Outline

1. Introduction: The uniqueness of cryptoassets for investments is explained. The methodology and aims of the thesis are stated and the organization of the thesis is presented.

2. Cryptoassets: This section provides a general overview of cryptoassets and crypto-markets.

3. Literature Review: Firstly, we discuss theories about the possible effect of liquidity on predictability. Subsequently, we delve into previous cryptoasset literature and the used economic approaches.

4. Methodology: The predictability and liquidity measures used in our analysis are delineated. Furthermore, we provide background for our economic approach.

5. Data: Data collection procedure is explained and summary statistics for the predictability measures and regression variables are computed.

6. Econometric Results: Results for cross-sectional and panel data analysis are presented. The results, limitations and possible topics for future studies are discussed.

7. Conclusion: The thesis is summarized, concentrating mainly on methodological approaches, results and future research directions.

Core bibliography

Alvarez-Ramirez, J., Rodriguez, E., & Ibarra-Valdez, C. (2018). Long-range correlations and asymmetry in the Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, 492, 948–955.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31–56.

Badev, A., & Chen, M. (2014). *Bitcoin: Technical Background and Data Analysis* (Finance and Economics Discussion Series No. 2014-104). Washington, D.C.: Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board.

Brauneis, A., & Mestel, R. (2018). Price discovery of cryptocurrencies: Bitcoinand beyond. *Economics Letters*, 165, 58–61.

Caporale, G. M., Gil-Alana, L., & Plastun, A. (2018). Persistence in the cryptocurrency market. *Research in International Business and Finance*, 46, 141–148.

Corwin, S. A., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance*, 67(2), 719–759.

European Central Bank Cryptoassets Task Force. (2019). *Cryptoassets: Implications for financial stability, monetary policy, and payments and market* *infrastructures* (European Central Bank Occasional Paper SeriesNo. 223). Frankfurt am Main: European Central Bank.

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92(2), 153–181.

Krištoufek, L. (2018). On Bitcoin markets (in)efficiency and its evolution. Physica A: Statistical Mechanics and its Applications, 503, 257–262.

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

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, 148, 80–82.

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

Author Supervisor

Chapter 1

Introduction

Cryptoassets can provide interesting and exciting investment opportunities. Although they are not yet widely accepted as currency (Mersch, 2018), cryptoassets may nevertheless pose as a unique asset class. From the turbulence on the crypto-markets, which have been the scene of year-on-year price rises even by a factor of 500 during early 2018 (European Central Bank Cryptoassets Task Force, 2019), to reports of artificially inflated volume figures (Singer, 2020), investing in cryptoassets may not be a simplistic matter, but it is definitely eventful. During the past years, crypto-markets have seen dramatic rises as well as periods of cooling down. At its peak in early 2018, the market capitalization of cryptoassets reached 650 billion EUR, falling to 96 billion EUR in the beginning of 2019 (European Central Bank Cryptoassets Task Force, 2019). Facing the volatile crypto-markets, investors may attempt to form various trading strategies to predict prices and exploit any trading opportunities while hoping to obtain profit. With this respect, it may be enlightening to examine the predictability of different cryptoassets and whether there are factors that can affect it.

Liquidity may then be an important aspect to consider. Avramov, Chordia, and Goyal (2006) point out that declines in market liquidity may lead to return predictability. Moreover, Chordia, Roll, and Subrahmanyam (2008) explain that liquid markets may have less predictable returns and may also be more efficient due to more arbitrage trading taking place. Hence, liquidity may have a negative impact on return predictability. This information may in turn be used by investors to help decide on and create trading strategies. Furthermore, previous literature mainly analyzes return predictability for cryptoassets. Only a few studies examine the relation between cryptoasset liquidity and predictability. Additionally, at the time of writing, we are not aware of research that would examine the effect of liquidity on cryptoasset predictability using panel data regressions. Hence, we analyze the relationship between cryptoasset liquidity and return predictability using both cross-sectional as well as panel data models.

We obtain data pertaining to 100 different cryptoassets for the years 2017 till 2019. This allows us to analyze a variety of different cryptoassets during an eventful time period for crypto-markets. We first calculate a number of predictability measures. Specifically, we include the Bartels test, runs test, Ljung-Box test, difference sign test, turning point test, two Lo-MacKinlay variance ratio tests and the Geweke and Porter-Hudak (GPH) estimate. Following the approach by Brauneis and Mestel (2018), we then construct a dependent variable, average cryptoasset rank, derived from the p-values of the respective tests as well as the absolute value of the GPH estimate. It may be noteworthy that all of the tests possess a null hypothesis suggesting a lack of predictability. We then build on the study by Brauneis and Mestel (2018) and construct three new dependent variables. We calculate the average p-values across the seven tests for each cryptoasset, the total number of tests which cannot reject the null hypothesis and a binary variable, equal to 1 when a cryptoasset has at least four tests for which the null hypothesis cannot be rejected.

Liquidity is measured by the Amihud illiquidity ratio and the Corwin-Schultz spread estimate. The Amihud illiquidity ratio is a commonly used measure for liquidity and is based on return and volume data (Amihud, 2002). The Corwin-Schultz spread estimates the bid-ask spread using only daily high and low asset prices (Corwin & Schultz, 2012) without relying on volume data. This may be a crucial factor to consider when examining cryptoassets, as there is some uncertainty about the validity and precision of cryptoasset volume data after reports of artificial volume inflation on crypto-markets (Singer, 2020). Following the approach by Brauneis and Mestel (2018), we also include other variables in our analysis. We specifically consider the logarithm of volume, turnover ratio and Garman-Klass volatility.

Subsequently, we estimate Ordinary Least Squares (OLS) and logit models on cross-sectional data. Furthermore, we extend our investigation and analyze the relationship between liquidity and predictability also using panel data. For panel data analysis, we estimate linear fixed effects models as well as the fixed effects logit model. Additionally, we conduct panel data analysis also for a subset of the full panel. The subset includes only 50 cryptoassets with the largest market capitalization values. We then examine whether the effect of liquidity for the cryptoassets with larger market capitalization is similar to the effect when considering the whole data set.

We begin by providing a general introduction to crypto-markets, describing cryptoassets and explaining some nuances of the market in Section 2. Section 3 presents a review of existing literature pertaining to cryptoassets and their liquidity and return predictability. We discuss theoretical background for the relationship between liquidity and predictability, introduce previous cryptoasset literature and outline the used methods. Section 4 describes the methodology used in our analysis. We delineate the used predictability measures and the subsequent construction of the dependent variables. We present the liquidity measures as well as the other independent variables and provide a theoretical background for our estimation methodology. In Section 5, data collection is described and summary statistics for the predictability measures as well as the regression variables are provided. Results of regressions on cross-sectional data, full panel data as well as subset panel data are presented in Section 6, followed by a discussion of the main findings and their limitations. Topics for future research are also proposed. Finally, Section 7 concludes the thesis.

Chapter 2

Cryptoassets

After the publication of the now well-known document by Nakamoto (2008), describing the mechanism behind Bitcoin, one of the most prominent cryptoassets took off and the crypto-markets have since developed considerably. According to the European Central Bank Cryptoassets Task Force (2019), in April 2013, there were 7 cryptoassets, while in May 2019 the number of cryptoassets reached about 1900. Currently, crypto-markets may pose as exciting investment opportunities. Yet as the European Central Bank Cryptoassets Task Force (2019) notes, cryptoassets have been linked with various drawbacks, such as money laundering, financing of terrorism and risks associated with consumer protection. Foley, Karlsen, and Putniņš (2019) also report that criminal activity may account for a startling 46% of all transactions that use Bitcoin. Crypto-markets may then be an eventful area for research.

Lacking a widely accepted definition, the European Central Bank (2019) highlights the absence of an underlying claim as well as the subsequent volatility and speculative nature as the distinguishing characteristics of cryptoassets. Additionally, cryptoassets are not yet recognized as a form of money. As Mersch (2018) explains, the use of cryptoassets as a medium of exchange is limited and significantly less widespread than that of more traditional currencies while the absence of an institution that would back cryptoassets also complicates using cryptoassets as a unit of account. Mersch (2018), however, claims that the largest factor causing the discrepancy between the common definition of money and the features of cryptoassets is the large volatility. Volatility makes cryptoassets difficult to use as a store of value, which in turn also worsens the position of cryptoassets as a unit of account (Mersch, 2018). Yermack (2013) reaches a similar conclusion, namely that Bitcoin[1] is not akin to typical currencies. Despite this assertion, Yermack (2013) does acknowledge that Bitcoin fulfills the medium of exchange requirement to a certain extent. The author, however, notes that the cryptoasset is still not widely used across the world to perform commercial transactions and thus labels Bitcoin as an investment viable for speculative purposes instead of a currency (Yermack, 2013). Therefore, although cryptoassets may not be accepted by economists as money, they may nevertheless be an interesting type of asset to investigate.

Cryptoassets are based on the Distributed Ledger Technology (DLT). Unlike traditional money, cryptoassets are decentralized (Schilling & Uhlig, 2019) and are thus also created in a decentralized method using the DLT. The DLT enables users to preserve their own copies of the ledger, enhancing data accessibility (Cryptoassets Task Force, 2018). Moreover, the DLT also implements cryptography and provides an opportunity for automated features[2] (Cryptoassets Task Force, 2018). Famously, the blockchain technology, a specific type of the DLT, is used for Bitcoin. Coins are created by mining, or solving mathematical problems (Schilling & Uhlig, 2019). As Badev and Chen (2014) expound, users verify transactions and obtain rewards in the form of new Bitcoins. Transactions are then stored as blocks recorded in a linear manner, or in other words, in a blockchain. According to Badev and Chen (2014), as more blocks are added to the growing chain, the former transactions get sealed and the likelihood of someone altering these transactions is verging on the impossible. Theoretically, a user would have to possess more than half of the total computing power of the entire network to attempt such a feat (Badev & Chen, 2014). The blockchain and the DLT thus possess numerous benefits, such as data accessibility due to the decentralized platforms, transparency in the form of recorded transactions as well as security of sealed transactions assured by the mechanism.

In order to purchase cryptoassets one would generally first need a wallet. Wallets are programs, secured by a private key, that allow a user to store and trade cryptoassets (Hileman & Rauchs, 2017). Numerous types of wallets are available. Some allow users to store multiple types of cryptoassets, while others are only suitable for a specific cryptoasset. Trading takes place on cryptoasset exchanges with some exchanges also providing wallet services. Exchanges may support trading cryptoassets for other cryptoassets or for fiat currencies. As

[1] Yermack (2013) focuses his study on Bitcoin specifically.

[2] For example, interest on a bond can be paid automatically (Cryptoassets Task Force, 2018).

an alternative to centralized crypto-exchanges, there exist peer-to-peer decentralized trading platforms, although Hileman and Rauchs (2017) mention that they are relatively less common. Crypto-exchanges may also charge deposit, trading and withdrawal fees.

Crypto-markets were an area of intense price developments especially in the past few years. At the time of writing, the cryptoasset market capitalization has reached its highest value at roughly 650 billion EUR in January 2018 after an increase during the years 2017–2018, according to the European Central Bank Cryptoassets Task Force (2019). After reaching its peak in early 2018, it then decreased and in the beginning of 2019, cryptoasset market capitalization was down to 96 billion EUR (European Central Bank Cryptoassets Task Force, 2019). Moreover, the surge in prices of cryptoassets in 2017–2018 was larger than that of historical bubbles. At its maximum in January 2018, the price of Bitcoin was roughly 19.5 times larger than during early 2017, while the prices of other cryptoassets increased by a staggering factor of 500 (European Central Bank Cryptoassets Task Force, 2019). These values are prominent even in comparison with historical bubbles of other assets. For example, the European Central Bank Cryptoassets Task Force (2019) mentions the dot-com bubble.

Interestingly, the European Central Bank Cryptoassets Task Force (2019) states that the total crypto-market capitalization correlates by 95% with price movements of Bitcoin. Moreover, it is pointed out that Bitcoin has lower volatility when compared with other cryptoassets. The European Central Bank Cryptoassets Task Force (2019) attributes this to a larger number of Bitcoin traders as well as the comparative maturity of Bitcoin. Furthermore, Bitcoin is one of the most dominant cryptoassets and its crypto-market share in early 2019 is declared to be 54% (European Central Bank Cryptoassets Task Force, 2019). This indicates the relative importance of Bitcoin within the crypto-markets.

As reported by the European Central Bank Cryptoassets Task Force (2019), the volatility of crypto-markets at the beginning of 2019 is still relatively large, despite being smaller than the volatility levels achieved during early 2018. Volatility hence remains a major drawback for cryptoassets. With this regard, stablecoins were introduced. Stablecoins aim to counter the high volatility of most cryptoassets and achieve a more constant value by having a more flexible supply or by being backed by other assets (European Central Bank Cryptoassets Task Force, 2019). This feature makes stablecoins quite unique and differentiates them from other cryptoassets. The European Central Bank Cryptoassets Task Force (2019) also adds that stablecoins are often used by traders on the crypto-markets as a hedging instrument. Stablecoins then present an interesting alternative to other cryptoassets. Moreover, considering that volatility may be considered to be one of the main hindrances preventing cryptoassets from being recognized as money (Mersch, 2018), it may be interesting to observe future developments of stablecoins. Nevertheless due to their unique characteristics and differences from other cryptoassets, researchers should beware of the treatment of stablecoins when analyzing crypto-markets.

In a consequential study, Schilling and Uhlig (2019) develop a model of an economy in which there exist both a traditional currency, the dollar, as well as a cryptoasset, the Bitcoin. The agents in this economy can either purchase and sell products using dollars or Bitcoins or avoid spending and keep the dollars or Bitcoins. The supply of the dollar is controlled by a central bank, while the supply of Bitcoin is decentralized. Schilling and Uhlig (2019) then derive a price bound for Bitcoin, which delineates when agents would use Bitcoin for speculation. Interestingly, Schilling and Uhlig (2019) also discover that according to the model, Bitcoin would follow a martingale process. This finding is distinctly interesting as it provides a theoretical setup for the development of Bitcoin prices with implications about the predictability of Bitcoin.

Cryptoasset exchanges have also attracted attention. Makarov and Schoar (2020) find that large price differences exist especially between exchanges in different countries, while relatively small price differences exist among different cryptoassets. Moreover, the authors also discover that arbitrage spreads among countries tend to move together (Makarov & Schoar, 2020). Arbitrage opportunities on various cryptoasset exchanges may be an interesting discovery, especially for cryptoasset traders. Yet, cryptoasset exchanges have been scrutinized also for different reasons. Singer (2020) reports that wash trading, which artificially inflates reported traded volumes for exchanges, may still be a recurring phenomenon on crypto-markets. Wash trading is a trading activity conducted in order to skew market information and may be carried out when the buyer and seller of an asset are the same entity. Moreover, crypto-exchanges may even have incentives to partake in wash trading, inflating their traded volumes, to seemingly improve their market activity in crypto-exchange rankings and entice new traders (Singer, 2020). Although websites, such as CoinMarketCap that aggregate market data for cryptoassets, are taking precautions and scrutinizing the data they obtain from crypto-exchanges (CoinMarketCap, 2020), complete elimination of wash trading practices is not a simple task. As Singer (2020) explains, to fully verify volumes reported by exchanges, the data aggregating websites would need sensitive information about the accounts conducting trades at the respective crypto-exchanges. Crypto-markets can thus be an interesting area for research. They involve novel technology, presented some dramatic market development in the past years, continue to spur discussion about the market characteristics and may provide an exciting, singular, albeit not necessarily simplistic, investment opportunity.

Chapter 3

Literature Review

3.1 Predictability, Liquidity and Market Efficiency

Predictability, liquidity and market efficiency are inter-linked concepts. While admitting that the definition of liquidity is not perfectly concrete, Avramov et al. (2006) consider liquidity to be the capacity for exercising sizable trades on a market while not triggering enormous price fluctuations. The authors, moreover, explain that return predictability may exist in the short-term in times of decreased market liquidity (Avramov et al., 2006). Chordia et al. (2008) also mention that timely and sizable arbitrage trading can result in decreased return predictability. Additionally, arbitrage trading would be more prevalent in liquid markets (Chordia et al., 2008). Following these notions, a negative relationship between liquidity and predictability would be indicated. Avramov et al. (2006) also find evidence in favor of this relation. The authors use stocks traded on the New York Stock Exchange during the years 1962–2002 and document a negative relationship between asset liquidity, measured by the Amihud illiquidity ratio, and return predictability in the form of short-term reversals (Avramov et al., 2006). Chordia et al. (2008) also document the relationship between liquidity and predictability. By analyzing stocks traded on the New York Stock Exchange during the 1993–2002 time period, the researchers discover that liquidity, measured by the bid-ask spread, negatively impacts return predictability (Chordia et al., 2008). Liquidity may, therefore, be an important aspect to consider with regards to return predictability. Moreover, there also exists evidence that liquidity may negatively impact return predictability. Some authors also propose a positive relationship between liquidity and market efficiency. According to Chordia et al. (2008), arbitrage trading, which is more prevalent in liquid markets, enhances price confluence to fundamental values, thus improving market efficiency. This notion is supported by Wei (2018), who also adds that there are more traders active on a liquid market. The positive relationship between liquidity and market efficiency may well hold, yet one should be careful not to equalize return predictability and market inefficiency without further examination.

Samuelson (1965) presents one of the first investigations of the concept of efficient markets and asset return predictability. A crucial implication of the examined model is that past price developments cannot be exploited to obtain profit (Samuelson, 1965). Market efficiency has since then been the subject of various studies and investigations. Jensen (1978) considers a market to be efficient when a trading strategy based on a specific information set cannot be exploited by traders to obtain abnormal profit. In one of the most seminal works in financial economics, Fama (1970) describes the weak, semi-strong and strong forms of market efficiency. In weak form efficiency, traders would not be able to gain abnormal profit by exploiting historical prices. For semi-strong efficiency all publicly available information, such as earnings announcements, are included in the considered information set. For strong form efficiency, traders would not be able to obtain additional profit even by applying private or insider information.

There is thus an important distinction between mere return predictability and market inefficiency. Timmermann and Granger (2004) clarify that return predictability does not have to immediately imply market inefficiency. Furthermore, asset prices are expected to follow a random walk in efficient markets only under specific conditions, which, for example, do not permit risk-premia or transaction costs (Timmermann & Granger, 2004). As Avramov et al. (2006) explain in their study, in spite of the presence of return predictability, once transaction costs are accounted for, trading strategies based on exploiting the presence of the short-term reversals are not profitable. Due to the absence of profitability as well as an adherence of the results to rational equilibrium models, the authors claim that market efficiency is in this case not profoundly violated (Avramov et al., 2006). It is then crucial to not only document predictability but also check for feasibility of profitable strategies before claiming market inefficiency. For example, serial correlation may suggest return predictability. Nevertheless, as Fama (1970) explains, it is convoluted to precisely identify a level of serial correlation that would enable a profitable trading strategy. Low levels of serial correlation, which can be detected, may not be enough for a successful profitable trading strategy. Additionally, there are also cases when the returns of an asset may have no serial correlation, yet a profit-making trading strategy based on nonlinear dependence may still be devised (Fama, 1970). Hence, Fama (1970) recommends to check the profitability of potential trading strategies. It is thus crucial to also check the feasibility of trading strategies before claiming market inefficiency. A simple detection of return predictability may not be enough to claim market inefficiency without further examination. For the purposes of our analysis, we will concentrate on the relationship between liquidity and predictability and leave market efficiency for future extensions and further studies.

3.2 Predictability and Liquidity of Cryptoassets

Various studies examine the predictability of cryptoassets, yet only few document the relationship between cryptoasset liquidity and return predictability. Previous literature varies in the types of crypto-markets analyzed, the considered time period as well as the final results.

Some studies focus only on specific cryptoassets. With this respect, Bitcoin is potentially the most favorite cryptoasset for analysis. Krištoufek (2018) analyzes the US dollar and Chinese yuan Bitcoin markets during the 2010–2017 time period, while Bariviera (2017) investigates Bitcoin for the years 2011–2017. Additionally, Urquhart (2016) examines Bitcoin during the years 2010–2016, using an aggregate, volume-weighted Bitcoin price index based on data from various Bitcoin exchanges. Moreover, while most studies include daily cryptoasset returns in their analyses, a notable exception is the research by Alvarez-Ramirez, Rodriguez, and Ibarra-Valdez (2018), who conduct an investigation of long-range properties of Bitcoin not only on a daily and hourly basis but even using second frequencies. The study is relatively unique in that it analyzes cryptoasset predictability at such depth. Alvarez-Ramirez et al. (2018) use daily prices from the years 2013–2017, hourly prices from May till June 2017 and high frequency prices from June 5^{th}, 2017 till June 6^{th}, 2017.

Other authors analyze more cryptoassets. For example, Caporale, Gil-Alana, and Plastun (2018) investigate persistence of four cryptoassets, namely Bitcoin, Litecoin, Ripple and Dash, during the years 2013–2017. In addition to these four cryptoassets, Zhang, Wang, Li, and Shen (2018) also consider Ethereum, NEM, Stellar, Monero and Verge. They use data from the 2013–2018 time period (Zhang et al., 2018). Krištoufek and Vošvrda (2019) study 6 cryptoassets from the beginning of 2015 till the end of June 2018 and 14 cryptoassets from the beginning of July 2017 till the end of June 2018. Additionally, Brauneis and Mestel (2018) examine 73 cryptoassets during the 2015–2017 time period. Furthermore, to the best of our knowledge, Wei (2018) presents one of the most comprehensive studies in terms of the number of examined cryptoassets. The author includes 456 different cryptoassets focusing on data from 2017 (Wei, 2018). Although only one year is used in the analysis, this study is nevertheless remarkable due to the large number of considered cryptoassets.

Multiple studies investigate the predictability of crypto-markets. As a disclaimer, a number of studies use predictability measures to also gauge market efficiency. Following our approach from Section 3.1, we refer to these measures only as measures for predictability. Zhang et al. (2018) as well as Wei (2018) document predictability for the examined cryptoassets. Additionally, Krištoufek (2018) reports predictability for both the US dollar and Chinese yuan Bitcoin markets. Krištoufek and Vošvrda (2019) also find predictability for Bitcoin, Dash, Litecoin, Stellar, Monero and Ripple. Moreover, studying only Bitcoin, Urquhart (2016) finds evidence of predictability for Bitcoin returns. Nevertheless, the researcher notes that there is some evidence for randomness of Bitcoin returns when considering only the second half of the sample, specifically for years 2013–2016 (Urquhart, 2016). This is an interesting finding and may indicate that Bitcoin returns are becoming less predictable over time. On this note, Brauneis and Mestel (2018) discover that Bitcoin is less predictable than 72 other cryptoassets.

Furthermore, Caporale et al. (2018) find evidence of persistence and thus also return predictability. Nevertheless, the authors observe changes in the Hurst exponent over time and claim that Bitcoin, Ripple, Dash and especially Litecoin are becoming progressively less predictable (Caporale et al., 2018). On a similar note, Bariviera (2017) find evidence of persistence for Bitcoin returns during the years 2011–2014 as well as evidence of decreases in the degree of predictability for the time period of 2014–2017. Additionally, Alvarez-Ramirez et al. (2018) discover time periods, when Bitcoin returns indicate a lack of predictability, as well as time periods, when Bitcoin returns could be anti-persistent. According to Alvarez-Ramirez et al. (2018), analyses with daily, hourly and even second frequency data support their findings. Hence, there exists a body of literature, documenting crypto-market predictability. Nevertheless, there may be some indications of potential decreases in the extent of return predictability and perhaps even a movement towards randomness of returns for some cryptoassets over time. It may thus be interesting to further document return predictability and any potential changes in its degree for these markets.

Although the predictability of crypto-markets has been documented before, only a few authors specifically include liquidity in their analyses. One of these researchers, Wei (2018), discovers that in comparison with the more liquid cryptoassets the less liquid ones are more predictable. Furthermore, Brauneis and Mestel (2018) also specifically examine the role of liquidity and find that the more liquid cryptoassets are less predictable. These results then suggest a negative relationship between cryptoasset liquidity and return predictability. Hence, although some research into the liquidity and predictability of crypto-markets has been conducted, it is not as of yet fully complete and further investigations of this topic may be enlightening.

3.3 Empirical Approaches

Various empirical approaches are used to gauge the predictability and liquidity of crypto-markets. Most studies include a wide array of tests to account for cryptoasset return predictability. Yet, only a minority of authors concretely analyze and measure liquidity in their analyses. Table 3.1 lists the measures for predictability and liquidity used in literature that deals with cryptoassets.

Most authors include a collection of tests to gauge the predictability of cryptoasset returns. For example, Wei (2018), Brauneis and Mestel (2018), Zhang et al. (2018) as well as Urquhart (2016) calculate the Ljung-Box test. One of the most widely included measures, the Ljung-Box test can be used to check for the presence of autocorrelation in cryptoasset returns. If detected, autocorrelation may indicate return predictability. Another test is the automatic Portmanteau test, performed by Zhang et al. (2018). This test may also be used to check for the presence of autocorrelation and is similar to the Ljung-Box test. Proposed by Escanciano and Lobato (2009), the automatic Portmanteau test is a version of the Box-Pierce test for autocorrelation. One of its main benefits is that the number of lags to be tested is calculated automatically and does not have to be chosen by the user (Escanciano & Lobato, 2009). Similarly as for the Ljung-Box test, the presence of autocorrelation would suggest a degree of predictability for cryptoasset returns.

Previous academic literature, pertaining to the predictability of cryptoassets, also includes several tests of randomness. Zhang et al. (2018), Urquhart

Table 3.1: Measures Used in Cryptoasset Literature

Predictability	Liquidity
Ljung-Box Test Automatic Portmanteau Test Bartels Test Runs Test Cox-Stuart Test Mann-Kendall Rank Test Turning Point Test Measure of Efficiency Brock-Dechert-Scheinkman Test Efficiency Index Variance Ratio Test Rank Variance Ratio Test Rank Score Variance Ratio Test Sign Variance Ratio Test Wild-bootstrapped (Automatic) Variance Ratio Test Hurst Exponent	Amihud Illiquidity Ratio Corwin-Schultz Spread Estimate

(2016), Wei (2018) as well as Brauneis and Mestel (2018) perform the Bartels rank test and the runs test for randomness. If randomness of cryptoasset returns cannot be rejected, a lack of predictability for the returns would be suggested. Furthermore, Zhang et al. (2018) also incorporate the Cox-Stuart test, the Mann-Kendall rank test and the turning point test. The Cox-Stuart test can be used to check for randomness of cryptoasset returns, while the Mann-Kendall rank test may be computed to detect the presence of positive or negative trends. The turning point test can be used to check whether cryptoasset returns are independent and identically distributed against the alternative hypothesis that the returns are not independent and identically distributed.

Moreover, in addition to numerous other tests, Brauneis and Mestel (2018) also include a relatively less common measure in their study, the so-called Measure of Efficiency. This measure is non-parametric, ranges from 0 to 1 and aims to gauge the performance of a passive trading strategy against an active one (Brauneis & Mestel, 2018). Another test is the Broock-Dechert-Scheinkman test, included in studies by Urquhart (2016), Wei (2018) as well as Brauneis and Mestel (2018). This non-parametric test can be used to check for independence of cryptoasset returns (Broock, Scheinkman, Dechert, & LeBaron, 1996). Furthermore, to investigate Bitcoin returns, Krištoufek (2018) uses the

Efficiency Index. This index, introduced by Krištoufek and Vošvrda (2014), can account for multiple measures of return predictability. For the study, Krištoufek (2018) specifically includes measures of long-range dependence, fractal dimension and approximate entropy. A similar approach is used in another study by Krištoufek and Vošvrda (2019).

Additionally, previous literature, dealing with the predictability of crypto-markets, presents a considerable variety of variance ratio tests. These include the standard variance ratio test, rank variance ratio test, rank score variance ratio test as well as the sign variance ratio test, all used by Zhang et al. (2018). Furthermore, Brauneis and Mestel (2018) use both the wild-bootstrapped variance ratio test as well as the wild-bootstrapped automatic variance ratio test in their analysis. The wild-bootstrapped automatic variance ratio test is also calculated by Zhang et al. (2018), Urquhart (2016) as well as Wei (2018). One benefit of the automatic versions of the variance ratio test is that necessary parameters are calculated automatically and do not require the user to specify them (Urquhart, 2016). Furthermore, according to Urquhart (2016), the wild-bootstrapped version of the automatic variance ratio test is claimed to possess superior performance for small samples, when compared to the standard automatic variance ratio test. Hence, as an alternative to the more classical variance ratio tests, the wild-bootstrapped automatic variance ratio test may present several benefits to a researcher. Furthermore, as Urquhart (2016) notes, variance ratio tests may be used to detect a random walk process. If cryptoasset returns were to follow a random walk, a lack of predictability would be suggested.

Many authors also investigate long-range dependence in crypto-market returns. For instance in their research, Zhang et al. (2018) include a considerably wide array of predictability measures. The authors, however, specifically concentrate on the generalized multi-fractal detrended fluctuation analysis to obtain the Hurst exponent. Using the Hurst exponent, researchers can check for long-range dependence. Alvarez-Ramirez et al. (2018) also use the detrended fluctuation analysis with a sliding window to calculate the Hurst exponent and analyze long-range properties of Bitcoin. Moreover, Urquhart (2016), Wei (2018) as well as Brauneis and Mestel (2018) also incorporate the Hurst exponent into their studies. The researchers calculate the Hurst exponent using the R/S analysis. Some authors, focusing on long-range dependence in the crypto-markets, use more than one method to calculate the Hurst exponent. Caporale et al. (2018) use two methods to examine persistence in crypto-markets, the

R/S analysis as well as the local Whittle estimator. Furthermore, the authors also study changes in the Hurst exponent over time (Caporale et al., 2018). Moreover, in addition to the R/S analysis, Bariviera (2017) also performs the detrended fluctuation analysis. The author claims that the R/S method is useful when checking for the presence of long-range dependence, while the detrended fluctuation analysis may be more suitable for examining changes that occur over time (Bariviera, 2017). Bariviera (2017) also states that the detrended fluctuation analysis may be preferred due to its robustness. Additionally, for the calculation of the Efficiency Index, Krištoufek (2018) as well as Krištoufek and Vošvrda (2019) also consider measures for long-range dependence, specifically the local Whittle and GPH estimators. Researchers, therefore, have a plethora of methods at their disposal to calculate the Hurst exponent and detect long-range dependence.

The predictability of cryptoasset returns has been measured in various ways by multiple researchers. Only a few authors, however, specifically include liquidity in their investigations. With this regard, several measures are used in cryptoasset literature to gauge liquidity. These measures are listed in Table 3.1. For example, Wei (2018) uses the Amihud illiquidity ratio, which is based on returns and daily traded volume figures. The author lists several benefits of this ratio, claiming that the Amihud illiquidity ratio is robust yet simple and may be calculated using daily data (Wei, 2018). Considering that the Amihud illiquidity ratio relies on reasonably easily obtainable data, it thus presents a significant benefit for researchers.

Nevertheless, one should take note that the Amihud illiquidity ratio depends on the reported traded volume. For cryptoassets this may pose as a unique caveat of the measure. As Singer (2020) writes, although websites reporting crypto-market data are taking precautions when reporting traded volumes, wash trading and volume inflation on crypto-exchanges may still be taking place. With this regard, it may be beneficial to not have to rely solely on a measure of liquidity that depends on the reported amount of volume. The Corwin-Schultz spread estimate is then a viable alternative. In addition to the Amihud illiquidity ratio, Brauneis and Mestel (2018) also include the Corwin-Schultz bid-ask spread estimate in their analysis to gauge cryptoasset liquidity. Furthermore, in a study dedicated to measuring price delays for cryptoassets, Köchling, Müller, and Posch (2019) also use both the Amihud illiquidity ratio as well as the Corwin-Schultz spread estimate to gauge liquidity of cryptoassets. As an alternative measure, the Corwin-Schultz bid-ask spread estimate uses only daily high and low prices in its calculation (Brauneis & Mestel, 2018). With this respect, it is uniquely useful in the context of measuring liquidity for cryptoassets.

After measuring liquidity, Wei (2018) as well as Brauneis and Mestel (2018) investigate its relationship with the predictability of cryptoasset returns. The authors use different approaches to quantify this relationship. Wei (2018) first calculates the Amihud illiquidity ratio to gauge liquidity. The author then separates the various cryptoassets into 5 groups, according to their liquidity. The predictability measures are then calculated separately for each group. More specifically, Wei (2018) includes the Ljung-Box test, runs test, Bartels test, Broock-Dechert-Scheinkman test, Hurst exponent as well as the wild-bootstrapped automatic variance ratio test. The researcher then reports the average p-values from the tests as well as the average value of the Hurst exponent for each cryptoasset group to find that the more liquid cryptoassets are also less predictable (Wei, 2018).

Brauneis and Mestel (2018) use a different method to investigate the predictability and liquidity of cryptoassets. The authors conduct a feasible generalized least squares regression, mentioning that this estimation may be preferred to the OLS regression when the dependent variable comprises estimates (Brauneis & Mestel, 2018). The independent variables are the time averages of the logarithm of market capitalization, logarithm of volume, Garman-Klass volatility, turnover ratio, Amihud illiquidity ratio and the Corwin-Schultz spread estimate, while the dependent variable is an average rank of predictability for the cryptoassets (Brauneis & Mestel, 2018). These ranks are based on the p-values of the Ljung-Box test, runs test, Bartels test, Broock-Dechert-Scheinkman test, variance ratio test, the wild-bootstrapped variance ratio test and the wild-bootstrapped automatic variance ratio test as well as the Measure of Efficiency value and the absolute difference from 0.5 of the Hurst exponent (Brauneis & Mestel, 2018). Moreover, Brauneis and Mestel (2018) note that the null hypotheses of all considered tests indicate a lack of predictability for cryptoasset returns. Hence, larger p-values from the tests would point to less return predictability for the cryptoassets. This approach then presents an interesting way to investigate the relationship between crypto-market predictability and liquidity. It includes a variety of different measures and does not rely on single proxies for liquidity and predictability. Moreover, it also applies a regression to examine the effect of liquidity on return predictability. Unfortunately, the time dimension is lost when averages in time are calculated. This, however, leaves room for further examination.

Previous literature quantifies the predictability of cryptoasset returns using numerous tests. Despite that, academic literature so far tends to concentrate almost exclusively on the predictability of crypto-markets with only a few studies that also account for the role of liquidity. Furthermore, to the best of our knowledge, at the time of writing no study of the relationship between cryptoasset liquidity and return predictability incorporates a panel data analysis. This topic may therefore be a particularly befitting area for research.

Chapter 4

Methodology

4.1 Predictability Measures

To examine return predictability of cryptoassets, we follow previous literature and select several predictability tests. In our analysis, we include 8 predictability measures, specifically the Ljung-Box test, Bartels rank test, runs test, turning point test, difference sign test, two Lo-MacKinlay variance ratio tests and the GPH estimate. The results of these measures are then aggregated into four main dependent variables, which are included in the econometric regressions. For clarification, returns are calculated as logarithmic returns for the entirety of our analysis.

Firstly, when considering the specific measures, we include the Ljung-Box test. Proposed by Ljung and Box (1978), the test is commonly used to check for autocorrelation. The test statistic is

$$Q = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{(T-k)} \tag{4.1}$$

where $\hat{\rho}_k$ is the sample autocorrelation at lag k and T is the number of observations (Ljung & Box, 1978). The null hypothesis states that the autocorrelations up to lag m are jointly zero. The test is often used in previous literature dealing with cryptoassets, for example in studies of the relationship between cryptoasset liquidity and return predictability by Brauneis and Mestel (2018) as well as Wei (2018). Moreover, it is also used by Urquhart (2016) and Zhang et al. (2018) when examining only cryptoasset predictability. For our purposes, the presence of autocorrelation would suggest a degree of predictability for cryptoasset returns.

Secondly, we include the Bartels rank test. Bartels (1982) presents a rank test for randomness based on the von Neumann ratio statistic. Each observation in a data sequence is given a rank and the test statistic is

$$RVN = \frac{\sum_{t=1}^{T-1}(R_t - R_{t+1})^2}{\sum_{t=1}^{T}(R_t - \bar{R})^2} \tag{4.2}$$

where R_t is the rank of observation t (Bartels, 1982). The inability to reject the null hypothesis of randomness for cryptoasset returns would indicate a lack of their predictability. The Bartels rank test is also used in previous studies by Brauneis and Mestel (2018), Wei (2018), Urquhart (2016) as well as Zhang et al. (2018).

Thirdly, we conduct the runs test for randomness. According to Gibbons and Chakraborti (2003), data is first classified into binary values for this test. If data is binary to begin with, no further transformation is needed. Otherwise, one may compare each observation in a data sequence to the median value and classify observations depending on whether they are larger or smaller than the median (Gibbons & Chakraborti, 2003). R is then the number of runs in a series with two types of elements (Gibbons & Chakraborti, 2003). The test statistic is

$$RS = \frac{R - \frac{2n_1n_2}{n_1+n_2} - 1}{\sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}} \tag{4.3}$$

where n_1 is the number of elements of one type and n_2 is the number of elements of the other type (Gibbons & Chakraborti, 2003). The null hypothesis states randomness of the data series. If the null hypothesis cannot be rejected, cryptoasset returns are suggested to be random and lack predictability.

Furthermore, we perform the turning point test of statistical independence. A turning point is said to occur at observation t, if either the previous observation $t-1$ and the following observation $t+1$ are both smaller than observation t or both observations $t-1$ and $t+1$ are larger than observation t (Brockwell & Davis, 2016). According to Brockwell and Davis (2016), the test statistic for the turning point test is

$$TP = \frac{P - \frac{2(T-2)}{3}}{\sqrt{\frac{16T-29}{90}}} \tag{4.4}$$

where P is the number of turning points and T is the number of observations. The null hypothesis states that the data is independent and identically distributed.

Moreover, we include two Lo-MacKinlay variance ratio tests. Introduced by Lo and MacKinlay (1987), this test checks for the presence of a random walk by observing variances of returns from different periods. According to Campbell, Lo, and MacKinlay (1997), the variance ratio can be written as

$$VR = \frac{\sigma_q^2}{\sigma_a^2} \tag{4.5}$$

where σ_q^2 is the variance of q-period returns and σ_a^2 is the variance of 1-period returns. Two test statistics may be calculated, one assuming homoskedasticity and one suitable for heteroskedasticity, to test for the null hypothesis of a random walk (Lo & MacKinlay, 1987). According to Campbell et al. (1997), the test statistic under homoskedasticity may be computed as

$$Z = \frac{\sqrt{nq} \times (VR - 1)}{\sqrt{\frac{2(2q-1)(q-1)}{3q}}} \tag{4.6}$$

while the test statistic that allows for heteroskedasticity can be written as

$$Z^* = \frac{\sqrt{nq} \times (VR - 1)}{\sqrt{4 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right)^2 \times \hat{\delta}_k}} \tag{4.7}$$

where

$$\hat{\delta}_k = \frac{nq \sum_{j=k+1}^{nq} (p_j - p_{j-1} - \hat{\mu})^2 (p_{j-k} - p_{j-k-1} - \hat{\mu})^2}{\left[\sum_{j=1}^{nq} (p_j - p_{j-1} - \hat{\mu})^2\right]^2} \tag{4.8}$$

for a series of log prices $\{p_0, p_1, ..., p_{nq}\}$ and the average of 1-period log returns $\hat{\mu}$ (Campbell et al., 1997). The presence of a random walk indicates a lack of predictability, hence if the null hypothesis cannot be rejected, we have evidence suggesting that cryptoasset returns are not predictable. For our analysis we conduct two Lo-MacKinlay variance ratio tests, one for a 2-day holding period and one for a 7-day holding period. Both of the tests are conducted with the test statistic that allows for heteroskedasticity. Variance ratio tests are often used in previous literature examining cryptoasset predictability. Brauneis and Mestel (2018), Urquhart (2016), Wei (2018) as well as Zhang et al. (2018) all incorporate versions of variance ratio tests. Some of the other types of variance ratio tests are the rank, rank score, sign, wild-bootstrapped and wild-bootstrapped automatic variance ratio tests.

Additionally, the difference sign test is calculated. This test, proposed by Moore and Wallis (1943), aims to check for randomness in a series using the positive or negative signs of first differences. The test statistic is

$$DS = \frac{S - \frac{T-1}{2}}{\sqrt{\frac{T+1}{12}}} \tag{4.9}$$

where S is the number of values for which the first difference of the data series is positive and T is the number of observations (Brockwell & Davis, 2016). The null hypothesis is randomness, while the alternative hypothesis is a positive or negative trend. If the null hypothesis cannot be rejected, a lack of predictability for cryptoasset returns would be suggested by the difference sign test.

The last of the included predictability measures is a measure for memory of a time series. Numerous previous studies include the Hurst exponent to gauge the predictability of cryptoasset returns. Although multiple authors use the Hurst exponent in their examination of cryptoasset returns, they also use different methods of calculation. Caporale et al. (2018) use the R/S analysis and the local Whittle estimator. Krištoufek (2018) as well as Krištoufek and Vošvrda (2019) include the local Whittle and GPH estimators. Bariviera (2017) uses the R/S analysis and the detrended fluctuation analysis. Additionally, the multi-fractal detrended fluctuation analysis is used by Alvarez-Ramirez et al. (2018) as well as Zhang et al. (2018). Brauneis and Mestel (2018), Urquhart (2016) and Wei (2018) use the R/S analysis. In our research, we opt for the Geweke and Porter-Hudak (GPH) estimator. Developed by Geweke and Porter-Hudak (1983), this method calculates the parameter d for a fractionally integrated process, $(1-B)^d X_t = u_t$, where u_t is stationary and linear. Geweke and Porter-Hudak (1983) use the logarithm of the spectral density function of X_t and compute

$$\begin{aligned}\log[I(\lambda_{j,T})] = &\log\left[\frac{\sigma^2 f_u(0)}{2\pi}\right] - d\log\left[4\sin^2\left(\frac{\lambda_{j,T}}{2}\right)\right] \\ &+ \log\left[\frac{f_u(\lambda_{j,T})}{f_u(0)}\right] + \log\left[\frac{I(\lambda_{j,T})}{f(\lambda_{j,T})}\right]\end{aligned} \tag{4.10}$$

where $I(\lambda_{j,T})$ is the periodogram at the harmonic ordinates, $\lambda_{j,T} = \frac{2\pi j}{T}$ for $j = 0, ..., T-1$, and f_u is the spectral density function of u_t. The parameter d

can then be estimated by OLS (Geweke & Porter-Hudak, 1983). Furthermore, d relates to the Hurst exponent, H, as $d = H - \frac{1}{2}$. Negative values of d indicate anti-persistence, while values of d that are larger than 0 suggest positive long-range dependence (Campbell et al., 1997). Estimating parameter d can then help detect whether cryptoasset returns follow a persistent or anti-persistent process and whether they possess a degree of predictability.

For the econometric analysis, we construct four dependent variables based on the described measures. Firstly, we follow Brauneis and Mestel (2018) and calculate the average ranks for the cryptoassets. The cryptoassets are ranked according to the results from the predictability measures. Lower ranks indicate less predictability, while higher ranks indicate larger predictability. Since all the tests have a null hypothesis that indicates a lack of predictability, the obtained p-values are used for the ranking. In the case of the GPH estimator, its absolute value is used, with larger values leading to higher ranks. In addition to the specification by Brauneis and Mestel (2018), we also calculate three new dependent variables. One of them is the average p-value for each cryptoasset across the 7 tests. The GPH estimator is not used in the calculation of this variable. Moreover, as all 7 tests have a null hypothesis that indicates a lack of predictability, lower average p-values would suggest more predictability. Another variable is the total number of tests, which indicate no predictability. In other words, for each cryptoasset we calculate the total number of tests, which had a p-value larger than 0.05. Finally, we also include a binary variable. This variable equals 1 if a cryptoasset has at least four tests for which the null hypothesis cannot be rejected. Otherwise, it takes on the value 0.

4.2 Liquidity Measures

There exist numerous measures for market liquidity, each with specific benefits as well as caveats. For our analysis we follow the approach of Brauneis and Mestel (2018) to calculate cryptoasset liquidity and consider two liquidity measures, the Amihud illiquidity ratio and the Corwin-Schultz spread estimate.

One of the liquidity measures is the Amihud illiquidity ratio. Proposed by Amihud (2002), the measure gauges illiquidity using the ratio of daily absolute returns to the daily traded dollar volume. Amihud (2002) describes it as a price impact measure, in other words, the response of prices per one dollar of traded volume. We calculate the Amihud illiquidity ratio for a cryptoasset as

$$\text{ami} = \frac{1}{T}\sum_{t=1}^{T}\frac{|r_t|}{vol_t} \tag{4.11}$$

where r_t are the returns and vol_t is the traded dollar volume at time t. Larger values for the Amihud illiquidity ratio indicate more illiquidity. Amihud (2002) mentions that there are other possible measures of liquidity, specifically the quoted or effective bid-ask spread. Yet, data to calculate these measures is not always readily available. The Amihud illiquidity ratio is then a viable alternative that allows a researcher to measure liquidity for longer periods of time for various financial assets. Wei (2018) also uses the Amihud illiquidity ratio to analyze the relationship between cryptoasset liquidity and predictability. Furthermore, Goyenko, Holden, and Trzcinka (2009) compare the performance of various measures of liquidity. They report that the Amihud illiquidity ratio performs well as a measure of price impact (Goyenko et al., 2009).

The Amihud illiquidity ratio may be a common, viable measure of liquidity, nevertheless in the case of cryptoassets one caveat prevails. The ratio depends on reported cryptoasset volume. Cryptoasset volume has been under scrutiny after reports of exchanges artificially inflating their data (Singer, 2020). Although websites that aggregate market data for cryptoassets, such as CoinMarketCap, aim to minimize the potential misreporting of volume data and exclude information from exchanges, which exhibit characteristics of artificial volume inflation (CoinMarketCap, 2020), wash trading practices and inflated volume figures have purportedly not fully disappeared from crypto-markets (Singer, 2020). Moreover, as Singer (2020) writes, countering wash trading practices and data misreporting is complicated, because in order to completely identify and deal with wash trading practices, data aggregating websites would need sensitive information about accounts trading at the respective exchanges. Considering that the Amihud illiquidity ratio depends on the reported traded volume, it may be of interest to also incorporate a measure of liquidity that would not include volume in its calculation.

With this regard, we compute the Corwin-Schultz spread estimate. Developed by Corwin and Schultz (2012) this measure estimates the bid-ask spread using daily high and low prices. Additionally, it is more accurate than other liquidity measures, namely Roll's estimator, which is another common measure of liquidity based on return covariances (Corwin & Schultz, 2012). The

Corwin-Schultz spread estimate for 2 days is defined as

$$S = \frac{2(e^{\alpha} - 1)}{1 + e^{\alpha}} \tag{4.12}$$

where

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}} \tag{4.13}$$

$$\gamma = \left[\log\left(\frac{H_{t,t+1}}{L_{t,t+1}}\right)\right]^2 \tag{4.14}$$

$$\beta = E\left\{\sum_{j=0}^{1}\left[\log\left(\frac{H_{t+j}}{L_{t+j}}\right)\right]^2\right\} \tag{4.15}$$

H_t and L_t are the daily high and low prices. $H_{t,t+1}$ and $L_{t,t+1}$ are respectively the high and low prices across days t and $t+1$. The Corwin-Schultz spread estimate may also be calculated for longer time frames. One should simply take the average of all overlapping 2-day spread estimates for the specified time period (Corwin & Schultz, 2012). Occasionally, the 2-day spread estimates may result in negative values. As Corwin and Schultz (2012) write, this may occur during particularly volatile time periods. For these cases, Corwin and Schultz (2012) recommend setting the spread estimates to 0 and note that this approach leads to more accurate spread estimates in comparison with spread estimates obtained by including or excluding the negative values. The Corwin-Schultz spread thus enables us to calculate a bid-ask spread estimate without the need for order book data. Moreover, this liquidity measure does not depend on traded volume, which may be of particular interest when analyzing cryptoassets. The inclusion of both the Amihud illiquidity ratio and the Corwin-Schultz spread estimate in the analysis may thus bring interesting insights.

Following the results of studies by Brauneis and Mestel (2018) as well as Wei (2018) we hypothesize that there will be a negative relationship between cryptoasset liquidity and return predictability. Larger values for the included liquidity measures indicate more illiquidity. Considering the return predictability variables, higher average cryptoasset ranks indicate more predictability, while larger average p-values indicate less predictability, as the null hypotheses of all included tests suggest a lack of predictability. Namely, we expect a positive effect of the liquidity measures on the dependent variable average cryptoasset rank. We also expect a negative impact of the liquidity measures on the average p-values, the number of tests suggesting a lack of predictability and the probability that a cryptoasset will have at least four such tests.

4.3 Econometric Approach

To the best of our knowledge only very few studies investigating the relationship between cryptoasset liquidity and predictability incorporate econometric models. Research by Brauneis and Mestel (2018) is one of the rare examples, which includes a cross-sectional regression. In our analysis, we initially follow the approach of Brauneis and Mestel (2018) and estimate a cross-sectional model with the average cryptoasset rank as the dependent variable. We then add onto previous research and estimate cross-sectional models with our other three dependent variables, namely the average p-value, the total number of tests which indicate a lack of predictability and a binary variable, equal to one if there are at least four such tests. Furthermore, we also estimate panel data models with analogous specifications. To the best of our knowledge, at the time of writing no other study uses panel data regressions to examine the relation between cryptoasset liquidity and return predictability.

All model specifications include as independent variables the two liquidity measures, the Amihud illiquidity ratio and the Corwin-Schultz spread estimate, as well as a set of control variables. Following the approach by Brauneis and Mestel (2018), the other independent variables include the Garman-Klass volatility measure, logarithm of traded dollar volume, logarithm of market capitalization and the turnover ratio. Logarithms of volume and market capitalization are rather straightforward measures and are meant to be directly calculated from obtained data. Proposed by Garman and Klass (1980), the Garman-Klass volatility estimator provides an efficient method of volatility estimation based on high and low as well as opening and closing prices. For a cryptoasset it is calculated as

$$\sigma = \sqrt{\frac{N}{n}\sum_{i=1}^{n}\left[\frac{1}{2}\left(\log\frac{H_i}{L_i}\right)^2 - (2\log 2 - 1)\left(\log\frac{C_i}{O_i}\right)^2\right]} \tag{4.16}$$

where H_i, L_i, C_i and O_i are the high, low, close and opening prices, respectively. N is the number of trading days in a year and n is the sample size. We also include the turnover ratio, computed as the traded dollar volume of a cryptoasset divided by its market capitalization.

For the first, cross-sectional part of our analysis we estimate three ordinary least squares (OLS) regressions and one logit model. Namely, we estimate the specifications

$$measure_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \tag{4.17}$$

$$P(bin_i = 1|\mathbf{x}_i) = \Lambda(\mathbf{x}_i'\boldsymbol{\beta}) \tag{4.18}$$

where $\mathbf{x}_i$ is a column vector of independent variables and Λ is the logistic function, $\Lambda(z) = \frac{e^z}{1+e^z}$ for real numbers z. The dependent variable *measure* is either the average cryptoasset rank, average p-value or the number of tests indicating a lack of predictability. The dependent variable *bin* equals 1 for cryptoassets, which have at least four tests indicating a lack of predictability.

Despite being a baseline regression model, OLS can be used in a multitude of various research scenarios and provides an efficient method for investigating relationships between variables. OLS relies on several assumptions, such as linearity in parameters, random sampling, no perfect collinearity among the regressors, zero mean for the error term and no correlation between the error terms and regressors, homoskedasticity of error terms and an asymptotically normal distribution of error terms (Wooldridge, 2002). To test for homoskedasticity of error terms the Breusch-Pagan test test may be used. White heteroskedasticity-robust standard errors may be used in the case heteroskedasticity is detected (Wooldridge, 2002). F-tests may be used to detect joint significance of variables. Moreover, the Variance Inflation Factor (VIF) can be used to gauge multicollinearity among the independent variables. Strong multicollinearity among the regressors may cause issues when estimating coefficients for these variables (Wooldridge, 2013). Although there is no fixed threshold which indicates that the degree of multicollinearity is too large, Wooldridge (2013) mentions that 10 is often used as a cutoff level. The R^2 and the adjusted R^2 may be used to evaluate OLS models.

The logit model may be used for regressions with binary dependent variables and can examine the impact of different variables on the probability of an event occurring. The logit model is nonlinear and is usually estimated by the maximum likelihood method. The Wald test and the likelihood ratio test may be used to test for joint variable significance (Wooldridge, 2013). Due to the non-linearity of the logit, average partial effects may be calculated. For the logit, partial effects can be written as

$$\frac{\partial P(bin_i = 1|\mathbf{x}_i)}{\partial x_{ik}} = \lambda(\mathbf{x}_i'\boldsymbol{\beta})\beta_k \tag{4.19}$$

where $\lambda(z) = \frac{e^z}{(1+e^z)^2}$ for real numbers z (Wooldridge, 2002). To obtain average partial effects one should calculate the marginal effects for every data point and then take the average of these values, although it is also possible to calculate marginal effects on the average (Greene, 2012). Moreover, due to the logistic function, the predicted values fall into the range between zero and one (Wooldridge, 2013). Unlike for OLS, the regular R^2 cannot be used to evaluate the logit. Other measures, such as the log-likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), may be used instead. The AIC and BIC may be calculated as

$$AIC = -2 \log L + 2k \tag{4.20}$$

$$BIC = -2 \log L + k \log n \tag{4.21}$$

where $\log L$ is the log likelihood, n is the number of observations and k is the number of estimated parameters (Greene, 2012). The two measures mainly differ in their preference for parsimonious models. When comparing information criteria, lower values indicate better performance among comparable models.

Alternatives to the logit model include the probit model and the linear probability model. The logit and probit models are similar, their main difference being that instead of the logistic function, the probit uses the standard normal cumulative distribution function. The logit, however, has several benefits when compared with the linear probability model. The linear probability model has in-built heteroskedasticity, although heteroskedasticity robust standard errors can be used to remedy this issue (Greene, 2012). Moreover, unlike the logit and probit, the linear probability model may predict probabilities that are larger than 1 or smaller than 0 (Greene, 2012). Considering these drawbacks, estimating the logit may be a more fitting approach.

In addition to the cross-sectional analysis, we also investigate the relationship between cryptoasset predictability and liquidity with panel data. For the panel data analysis, we include the same set of independent as well as dependent variables as for the cross-sectional case. Instead of calculating each of the variables across the whole time period, we use a rolling window of 180 days, approximately half a year, with a time skip of 7 days. Panel data analysis thus allows us to not only investigate the relationship between cryptoasset predictability and liquidity across various cryptoassets but also across time.

When dealing with panel data, the pooled, random effects and fixed effects models are some of the main regressions that may be estimated. The baseline panel data model, the pooled model, can be used when heterogeneity does not vary for different individuals (Greene, 2012). It can, furthermore, be estimated by OLS. Yet, there may be cases when heterogeneity varies among the individuals and other panel data models should be used. The fixed effects model should be used when the unobserved individual effects are correlated with the independent variables (Greene, 2012). In addition to individual fixed effects, time effects may also be added. Yet, despite the usefulness of the fixed effects model, as Greene (2012) notes, the model does not estimate coefficients for variables, which do not vary across time. Another common panel data model is the random effects model. It may be used when the unobserved individual effects are uncorrelated with the independent variables. As Greene (2012) explains, this may be the case when the individual units in the model come from a large population.

Several tests may be used to choose from the three panel data models. According to Greene (2012) Hausman's specification test can be used to choose between the fixed effects and random effects models. Under the null hypothesis, both models are consistent, but the random effects model is also efficient, while the alternative hypothesis states that only the fixed effects model is consistent. Moreover, F-tests may be used to check for individual and time effects. These tests may be used to decide between the pooled or fixed effects models. We conduct the Hausman specification test as well as the F-tests for individual and time effects and identify the need to use the individual and time fixed effects model. We thus estimate fixed effects models based on analogous specifications as for the cross-sectional case. More precisely, we estimate

$$measure_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \epsilon_{it} \tag{4.22}$$

$$P(bin_{it} = 1|\mathbf{x}_{it}) = \Lambda(\alpha_i + \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta}) \tag{4.23}$$

where Λ is the logistic function, $\mathbf{x}_{it}$ is the column vector of independent variables, α_i are individual effects and γ_t are time effects. The variable *bin* equals one when there are at least four tests indicating a lack of predictability and *measure* is either the dependent variable cryptoasset rank, average p-value or the number of tests indicating no predictability. The fixed effects model also relies on several assumptions, such as linearity in parameters, random sampling on the cross-sectional level, no perfect collinearity and regressors varying in time, strict exogeneity, homoskedasticity, uncorrelated idiosyncratic errors and asymptotically normally distributed idiosyncratic error terms (Wooldridge, 2013). Clustered standard errors on the individual or time level may be used.

Equation 4.23 presents the fixed effects logit model. The fixed effects logit can be estimated by the conditional logit estimator. According to Stammann, Heiss, and McFadden (2016), this estimator is consistent as the number of individuals approaches infinity for a fixed number of time segments. Yet, as the authors note, the conditional estimator does not provide the estimates of the fixed effects, which then means that partial effects cannot be computed. Additionally, this approach can be computationally challenging for larger panel data sets, especially for those with a larger number of time periods (Stammann et al., 2016). As an alternative, Stammann et al. (2016) propose an unconditional estimator. The authors incorporate a pseudo-demeaning algorithm together with a correction of bias and develop an estimator that not only possesses properties corresponding to those of the conditional estimator but also has improved computational efficiency, particularly for larger panel data sets. Moreover, Stammann et al. (2016) provide this estimator in an R package. The package, furthermore, allows users to calculate average partial effects for fixed effects logit models, although the calculation for interaction or squared terms is not yet supported. This approach then possesses significant advantages in estimating fixed effects logit models for larger panel data sets and greatly aids us in our endeavors.

Chapter 5

Data

5.1 Data Collection

We collect data for 100 largest cryptoassets, based on their market capitalization, that have market data from the beginning of January 2017 till the end of December 2019. Our obtained data set thus consists of daily market data for 100 cryptoassets across a time period of three years. This allows us to investigate a sizable number of cryptoassets during the 2017 boom as well as the subsequent fall on the crypto-markets. Table A1 in the Appendix lists all of the cryptoassets used in the analysis. Both tokens as well as coins are included. Our data set, however, does not include stablecoins. Stablecoins differ from other cryptoassets as they aim to have lower volatility and may be backed by other assets (European Central Bank Cryptoassets Task Force, 2019). For example, Tether is a stablecoin supposedly backed by the US dollar. Its aim is to maintain a stable price equal to 1 USD. USD Coin is another stablecoin which tries to maintain its peg to the US dollar. Their prices thus do not vary much so that predictions are very limited.

Cryptoasset data is obtained from CoinMarketCap. We gather data pertaining to daily opening, closing, high and low prices as well as market capitalization and volume. All variables are in US dollars. CoinMarketCap, a well-known website for cryptoasset information, provides market data for a multitude of cryptoassets across exchanges. CoinMarketCap (2020) calculates prices as volume weighted averages. Reported volume is the total daily volume across all exchanges for a given cryptoasset. Reported market capitalization is computed as a multiplication of the price of a given cryptoasset by its circulating supply (CoinMarketCap, 2020). Moreover, information from some exchanges is excluded from the calculations. This is only in cases when the exchanges exhibit characteristics akin to artificially inflating the reported trading volumes (CoinMarketCap, 2020). Nevertheless, as Singer (2020) writes that wash trading and thus inflated market data may still occur on the crypto-markets, reported volumes should not be completely free from scrutiny.

5.2 Summary Statistics

To visualize the results of the predictability tests, Table 5.1 provides the summary statistics for the predictability measures. There are eight measures in total, namely the seven tests and the GPH estimate. For the tests, we report the summary statistics using the computed p-values. Lo-MacKinlay variance ratio tests (2) and (7) refer to the two conducted tests, one using a 2-day holding period and one for a 7-day holding period, respectively. The GPH estimate is presented in the table without any further transformations.

Table 5.1: Summary Statistics for Predictability Measures

Measure	Mean	St. Dev.	Minimum	Maximum	Median
Bartels	0.135	0.228	0.000	0.972	0.008
Runs	0.118	0.192	0.000	0.952	0.027
Turning Point	0.299	0.310	0.000	0.962	0.177
Difference Sign	0.489	0.278	0.005	1.000	0.464
Ljung-Box	0.056	0.119	0.000	0.656	0.000
GPH estimate	0.157	0.140	−0.189	0.441	0.162
Lo-MacKinlay (2)	0.279	0.323	0.000	0.966	0.129
Lo-MacKinlay (7)	0.258	0.281	0.000	0.976	0.148

All of the tests have mean p-values larger than 0.05, the smallest being 0.056 for the Ljung-Box test. On the other hand, when considering the median values, three tests, specifically the Bartels test, runs test and the Ljung-Box test, have median p-values below 0.05. Furthermore, 74 cryptoassets have p-values lower than 0.05 for the Ljung-Box test. This is the largest number of cryptoassets, for which the null hypothesis is rejected, among all of the tests. In contrast, the difference sign test has only 7 cryptoassets, for which the null hypothesis is rejected. 63 cryptoassets have p-values lower than 0.05 for the Bartels test and 52 cryptoassets have p-values below 0.05 for the runs test. For the turning point test, this number is 29. For the Lo-MacKinlay variance ratio tests (2) and (7) the hypothesis of a random walk is rejected for 44 and

31 cryptoassets, respectively. MintCoin, Pandacoin and Swing are among the cryptoassets with the lowest p-values across all of the considered tests, while Bitcoin, Dash and Burst are among the cryptoassets with the largest p-values.

The GPH estimate may range from −0.5 to 0.5. As can be seen from Table 5.1, for the cryptoasset data set it reaches a minimum value of −0.189 and a maximum of 0.441. Notably, Swing, Pandacoin and MintCoin are among the cryptoassets with the most negative values for the GPH estimate, suggesting anti-persistence. The maximum GPH estimate is obtained by DigitalNote, while the estimates closest to 0 are for Clams and I/O Coin. Moreover, both the mean and median values of the GPH estimate are positive.

Table 5.2: Summary Statistics

Variable	Mean	St. Dev.	Minimum	Maximum	Median
rank	50.728	16.743	21.000	93.000	48.375
average p-value	0.233	0.139	0.001	0.659	0.202
total tests	3.980	2.015	0.000	7.000	4.000
ami	0.027	0.166	0.000	1.560	0.000
spread	0.181	0.112	0.040	0.569	0.146
gk	2.230	0.957	0.747	6.676	2.101
logvol	11.923	3.625	4.074	22.203	11.656
turnover	0.045	0.061	0.001	0.355	0.026
logMC	16.711	2.475	11.662	25.199	16.275

Note: Amihud illiquidity ratio (ami), Corwin-Schultz spread estimate (spread), Garman-Klass volatility (gk), log volume (logvol), turnover ratio (turnover), log market capitalization (logMC)

Summary statistics for variables that are used in the econometric analysis are presented in Table 5.2. The first three variables shown in the table are the three dependent variables, specifically the average cryptoasset rank, average p-value and the total number of tests which suggest the lack of predictability. Lower cryptoasset ranks indicate less predictability. The mean rank is 50.728, while the median is 48.375. The mean p-value is 0.233 and the median is 0.202. The mean number of tests, which could not reject the lack of predictability, is about 3.98 with the median being 4. Bitcoin obtained both the minimum cryptoasset rank of 21 as well as the maximum average p-value of 0.659. Conversely, Pandacoin obtained the maximum cryptoasset rank of 93 and the minimum average p-value of roughly 0.001. Out of the seven considered tests, no tests indicated the lack of predictability for Pandacoin with 0 being the minimum value for this variable. All of the tests suggested no predictability for Bitcoin, Burst, Clams, CPChain, Dash, Feathercoin, Groestlcoin, LBRY Credits, NavCoin, Syscoin, Viacoin and Zcash. Looking at these predictability variables, the cryptoassets Bitcoin and Dash are most notably among the least predictable, while Pandacoin, Digitalcoin, Swing and Mintcoin are among the most predictable. Furthermore, 54 of the 100 cryptoassets had at least four tests that indicated a lack of predictability.

Two liquidity measures are considered, the Corwin-Schultz spread estimate and the Amihud illiquidity ratio. For both variables, larger values indicate less liquidity. The average value for the Corwin-Schultz spread is about 0.181, while the median is roughly 0.146. The average Amihud illiquidity ratio is 0.027 and the median is roughly 0. Considering the data, the cryptoassets Bitcoin, Dash, Litecoin, Ethereum and Monero are among the more liquid ones, while Pandacoin, Digitalcoin and Mintcoin are among the least liquid. Other variables include the Garman-Klass volatility measure, logarithm of volume, logarithm of market capitalization as well as the turnover ratio. Bitcoin has the lowest Garman-Klass volatility of 0.747 as well as the largest log volume of 22.203 and log market capitalization of 25.199. At 6.676 Pandacoin has the largest Garman-Klass volatility. Swing has the smallest log volume of 4.074 as well as log market capitalization of 11.662. Ethereum Classic has the largest turnover ratio, about 0.355, while Ixcoin has the smallest one, 0.001.

Chapter 6

Econometric Results

6.1 Results for Cross-Section Models

The relationship between cryptoasset liquidity and return predictability is first investigated using cross-sectional data. We consider four different dependent variables, the average rank of a cryptoasset based on the predictability measures, the average p-value from the predictability tests, the total number of tests for which the lack of predictability could not be rejected and a binary variable equal to one if at least four tests indicated no predictability. The independent variables include Garman-Klass volatility (gk), logarithm of volume (logvol), turnover ratio (turnover), logarithm of market capitalization (logMC) as well as the two liquidity measures, the Amihud illiquidity ratio (ami) and the Corwin-Schultz spread estimate (spread). At first we include all of the variables in the regressions, but upon further inspection, multicollinearity is detected between volume and market capitalization using the VIF with values well above 10. Considering the VIF values as well as model performance, specifications including only the variable volume are chosen. Baseline regressions for all four dependent variables are presented in Table 6.1. The average partial effects for the logit model are shown in Table A2 in the Appendix. Upon detection of heteroskedasticity in the error terms, White heteroskedasticity-robust standard errors are used.

All of the models indicate a significant, negative relationship between return predictability and liquidity of cryptoassets as measured by the Corwin-Schultz spread estimate. Larger values of the Corwin-Schultz spread estimate point towards lower liquidity. From the regression using the average rank as the dependent variable, an increase in the spread estimate by 0.1 would lead to the

Table 6.1: Baseline Cross-Section Models

	OLS			Logit
	rank	average p-value	total tests	
Constant	50.087***	0.068	3.155*	−2.849
	(11.138)	(0.086)	(1.257)	(3.110)
gk	−1.754	0.015	0.564	1.374
	(3.286)	(0.021)	(0.383)	(1.011)
ami	9.356	0.004	−1.144*	−25.781
	(9.267)	(0.034)	(0.544)	(71.586)
spread	83.542***	−0.391*	−11.867***	−19.142**
	(23.896)	(0.157)	(2.365)	(6.601)
logvol	−0.840	0.017**	0.141	0.285
	(0.641)	(0.006)	(0.071)	(0.167)
turnover	−17.930	−0.054	1.534	−0.654
	(27.053)	(0.267)	(4.419)	(6.865)
R^2	0.438	0.359	0.449	
Adjusted R^2	0.408	0.325	0.420	
F Statistic	14.671***	10.542***	15.335***	
Log Likelihood				−45.634
AIC				103.268
BIC				118.899

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Standard errors are in parentheses.

average rank of a cryptoasset increasing by roughly 8, other variables remaining constant. Considering that larger ranks are associated with more predictability, we find a relation between lower cryptoasset liquidity and larger predictability. The second model finds a significant, negative effect of the Corwin-Schultz spread on the average p-value from the predictability tests. All of the considered tests have no predictability as the null hypothesis, larger p-values would thus indicate less predictability. We also find a significant, negative impact of lower liquidity on the total number of tests for which the lack of predictability cannot be rejected. Moreover, the spread estimate negatively affects the probability of a cryptoasset having at least four tests that indicate no predictability. The average partial effect of the Corwin-Schultz spread estimate is also found to be significant and negative, as exhibited in Table A2 in the Appendix. None of the other average partial effects are significant.

Interestingly, the relationship between predictability and liquidity is found to be significant mainly for the Corwin-Schultz spread. The Amihud illiquidity ratio is significant only in the model with total tests as the dependent variable.

In this case, its sign supports the hypothesis of a negative relationship between cryptoasset liquidity and return predictability, in accord with the results for the Corwin-Schultz spread. The Amihud illiquidity ratio is, furthermore, insignificant in all of the other models. Its average partial effect is also insignificant. Cryptoasset volume is significant only in the model with the average p-values as the dependent variable. The sign of its coefficient suggests that cryptoassets with larger volumes obtain larger average p-values and are less predictable. Other variables remain insignificant.

Additionally, we investigate whether there exists a quadratic relationship between cryptoasset liquidity and predictability. Perhaps there exists a turning point and liquidity may have a different effect on cryptoasset predictability once it passes beyond a certain level. For example, increases in liquidity may have a negative effect on predictability for already quite liquid cryptoassets, as opposed to the impact on the predictability of more illiquid cryptoassets. Table A3 in the Appendix reports the results for analogous regressions, which include squared terms for all variables. Using the F-test, Wald test and likelihood ratio test to check for joint significance, we consider both individual as well as joint significance of variables and construct the models presented in Table 6.2. The average partial effects from the logit models from Tables A3 and 6.2 are shown in Table A2 in the Appendix.

Table 6.2: Cross-Section Models

	OLS			Logit
	rank	average p-value	total tests	
Constant	34.779***	0.337***	5.885***	3.650***
	(2.333)	(0.022)	(0.268)	(0.833)
spread	88.169***	−0.574***	−10.533***	−19.543***
	(11.030)	(0.078)	(0.994)	(4.478)
turnover	−24.841*	0.243***	3.115**	6.334**
	(12.282)	(0.071)	(1.069)	(2.406)
turnover^2	44.657***	−0.474***	−6.343***	−13.915***
	(12.116)	(0.086)	(0.925)	(3.519)
R^2	0.494	0.410	0.524	
Adjusted R^2	0.479	0.391	0.509	
F Statistic	31.293***	22.222***	35.229***	
Log Likelihood				−38.054
AIC				84.108
BIC				94.528

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Standard errors are in parentheses.

Interestingly, the turnover ratio is found to have a significant quadratic relationship with cryptoasset predictability. Its squared component is significant in all model specifications. For initial values of the turnover ratio, increases in the ratio lead to decreases in the average rank. As the turnover ratio further increases, this relationship changes and for larger values of the turnover ratio, increases in the ratio result in higher average ranks. The relationship for the other dependent variables is converse. For example, for the initial values of the turnover ratio, the average p-value increases as the ratio increases. As the turnover ratio increases further, the relationship changes and the average p-values eventually decrease. There is thus evidence that the degree of predictability at first decreases, but eventually starts increasing as the turnover ratio increases. Moreover, the average partial effect of the turnover ratio, shown in Table A2, is also significant with a positive sign. Additionally, we do not find evidence for a quadratic effect of volume and Garman-Klass volatility on cryptoasset predictability.

The Corwin-Schultz spread estimate is highly significant for all specifications and possesses the same signs as for the baseline models, thus having the same direction of effect. Furthermore, we do not find evidence in favor of a quadratic relationship between cryptoasset liquidity and return predictability. Only the linear component of the Corwin-Schultz spread estimate is significant, while both the linear and quadratic terms of the Amihud illiquidity ratio are individually as well as jointly insignificant across all four model specifications. We thus have evidence suggesting a linear, negative relationship between cryptoasset liquidity, measured by the Corwin-Schultz spread, and return predictability.

6.2 Results for Panel Data

To further investigate the relationship between liquidity and predictability of cryptoassets we estimate panel data models. The set of independent and dependent variables is the same as for the regressions on cross-sections. We test the pooled, random effects and fixed effects models using the Hausman specification test as well as the F-test for individual and time effects and conclude the need to use the individual and time fixed effects model. As in the case with cross-section models, multicollinearity between the variables volume and market capitalization is indicated by the VIF. Following the former approach, only models with the variable volume are reported. The results for the baseline linear fixed effects and the logistic fixed effects models using the full data set are presented in Table 6.3. The average partial effects for the fixed effects logit model are displayed in Table A2 in the Appendix. Clustered standard errors on the individual and time level are used for all panel data models.

Table 6.3: Baseline Panel Data Models

		Linear FE		Logit FE
	rank	average p-value	total tests	
gk	−0.601	0.002	0.329	1.093**
	(2.191)	(0.026)	(0.218)	(0.416)
ami	−1.122	0.026**	0.245	0.313
	(0.727)	(0.009)	(0.142)	(0.447)
spread	−20.778	0.165	4.140**	12.697***
	(12.733)	(0.158)	(1.302)	(3.530)
logvol	−1.803***	0.018***	0.233***	0.474**
	(0.423)	(0.005)	(0.051)	(0.171)
turnover	0.462	0.014	−0.694	−0.677
	(6.425)	(0.080)	(0.533)	(3.792)
Overall R^2	0.409	0.410	0.519	
Within R^2	0.051	0.143	0.194	
F-Statistic	7.768***	7.537***	9.417***	
Log Likelihood				−1839.279
AIC				3688.558
BIC				3725.960

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Clustered standard errors are in parentheses.

Interestingly, the Amihud illiquidity ratio is significant in the model with average p-values as the dependent variable. Furthermore, the Corwin-Schultz spread estimate is significant in the fixed effects logit and in the regression with total tests as the dependent variable. All three coefficients are positive, indicating a positive relation between liquidity and predictability for cryptoassets. Moreover, the average partial effect of the spread estimate, reported in Table A2 in the Appendix, is also found to be significant with a positive coefficient. This result is somewhat counter-intuitive, as one would expect that more liquid cryptoassets would also be less predictable. The finding is also contrary to the results from cross-sectional data, which suggest a negative relationship between cryptoasset liquidity and return predictability.

Only volume is found to have a significant relation with the predictability of cryptoasset returns across all of the specifications. It negatively impacts the average rank of cryptoassets and positively affects the average p-values, the number of tests for which the lack of predictability cannot be rejected and the probability that there are at least four such tests. The average partial effect of the logarithm of volume, provided in Table A2, is also significant and positive. Volume is then found to have a negative impact on the degree of cryptoasset predictability. This is an interesting result and indicates that cryptoassets, which are traded more, are also less predictable. Moreover, Garman-Klass volatility is found to have a positive effect on the probability of a cryptoasset having at least four tests that indicate no predictability. Its average partial effect is also significant and positive. The turnover ratio remains individually insignificant in all four models.

We thus find some evidence of a positive relationship between cryptoasset liquidity and return predictability from baseline panel data regressions. This finding is, however, contrary to the results from the cross-sectional analysis. Furthermore, although the overall R^2 of the fixed effects linear regressions is around 0.4-0.5, the within R^2 is quite low for all three specifications. The within R^2 measures how well the independent variables explain variation in the dependent variable within each cryptoasset across time. Considering the values, it seems that variation in return predictability across time within the cryptoassets is not fully captured by the regressors. Following the same procedure as in the cross-sectional analysis, we further examine whether a potential quadratic relationship between cryptoasset liquidity and predictability exists. The results for regressions with squared terms for all independent variables are shown in Table A4 in the Appendix. Joint significance of regressors is tested using the F-tests, Wald tests and likelihood ratio tests. We then construct models based on joint and individual significance and present them in Table 6.4.

There is not much evidence for a quadratic effect of liquidity on cryptoasset predictability. Only the fixed effects logit has quadratic terms for the Amihud illiquidity ratio and the Corwin-Schultz spread estimate. Moreover, while the squared component of the Amihud illiquidity ratio is also individually significant, the squared term of the Corwin-Schultz spread is only jointly significant. Note that the fixed effects logit model is the full model with all squared terms, as all the variables are determined to be jointly significant by the likelihood ratio test. When considering the linear terms of the liquidity measures, the Amihud illiquidity ratio has an individually significant, positive effect on the average p-value, similarly as for the baseline regressions. Additionally, the linear components of the Corwin-Schultz spread estimate are significant in the fixed effects logit and in the model that uses the total number of tests as the dependent variable. Both coefficients are positive, suggesting a positive relationship between cryptoasset liquidity and return predictability.

Table 6.4: Panel Data Models

	Linear FE rank	 average p-value	 total tests	Logit FE
gk	321.006**	−2.987*	43.569	123.992**
	(114.288)	(1.260)	(26.199)	(42.796)
gk^2	−170.418**	1.372*	17.962**	25.320
	(55.687)	(0.614)	(5.598)	(15.970)
ami	−1.207	0.027**	0.289	−6.932
	(0.710)	(0.008)	(0.154)	(11.507)
ami^2				18.191*
				(8.217)
spread			4.317**	187.216***
			(1.314)	(51.032)
$spread^2$				0.490
				(14.352)
logvol	−1.684***	0.018***	0.204***	209.554***
	(0.385)	(0.004)	(0.045)	(52.774)
$logvol^2$				91.594**
				(28.236)
turnover				−81.470
				(63.969)
$turnover^2$				−76.612
				(68.304)
Overall R^2	0.410	0.410	0.524	
Within R^2	0.052	0.143	0.201	
F-Statistic	10.903***	10.239***	9.594***	
Log Likelihood				−1733.394
AIC				3486.788
BIC				3561.592

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Clustered standard errors are in parentheses.

The models in Table 6.4, however, consistently document the effect of volume. For all specifications, the linear term of the logarithm of volume is highly significant. Moreover, when considering the fixed effects logit model, even its squared component is found to be significant with a positive coefficient. The logarithm of volume positively impacts the average p-values and the total number of tests for which the lack of predictability cannot be rejected, while negatively affecting the average rank. These results suggest that there is a significant, negative relationship between cryptoasset volume and the degree of return predictability. Although this finding contradicts the results of the cross-sectional analysis, it is consistent with the baseline panel data regressions. One may also, however, note that similarly to the case with the baseline panel data regressions, the within R^2 remains quite low.

Considering other variables, Garman-Klass volatility and its square component are jointly significant in all four models. For regressions with average rank and average p-value as the dependent variables, the degree of return predictability initially increases for rising Garman-Klass volatility. After the turning points, return predictability starts decreasing for larger values of the volatility measure. For the model with total tests as the dependent variable, increases in Garman-Klass volatility lead to decreases in return predictability. In the fixed effects logit, only the linear component of volatility is individually significant. With this respect, one may also note that although there is some evidence of a quadratic relationship between Garman-Klass volatility and cryptoasset predictability for panel data, these results are not supported by cross-sectional regressions. Moreover, although there is evidence of a quadratic relation between the turnover ratio and cryptoasset predictability for the cross-sectional case, panel data does not provide much evidence in favor of this relationship.

As an extension, we conduct the same panel data regressions using a subset of the data, containing only the top 50 cryptoassets as measured by their average market capitalization over the entire considered time period. The full data set contains more frequently traded cryptoassets with large market capitalization as well as more niche cryptoassets, which are traded comparatively less. The relationship between predictability and liquidity may then be different for the cryptoassets with larger market capitalization values. The results of baseline regressions are displayed in Table 6.5. The Amihud illiquidity ratios are multiplied by 1000 for clarity in the subset regressions.

The only individually significant variable in these regressions is the logarithm of volume. It negatively affects the average rank and positively impacts the total number of tests, indicating a lack of predictability, as well as the probability that a cryptoasset has at least four such tests. Hence, there is some evidence for a negative effect of cryptoasset volume on the degree of return predictability. All other variables remain individually insignificant. Moreover, the average partial effects from the fixed effects logit, exhibited in Table A2 in the Appendix, are also all insignificant. Following previous approaches,

Table 6.5: Baseline Panel Data Models (Subset)

	Linear FE			Logit FE
	rank	average p-value	total tests	
gk	3.426	−0.037	−0.168	−0.693
	(3.182)	(0.037)	(0.194)	(1.165)
ami	7.831	−0.084	−0.954	0.412
	(5.910)	(0.061)	(0.544)	(2.663)
spread	3.311	−0.056	1.023	5.817
	(21.970)	(0.245)	(1.515)	(10.974)
logvol	−1.934*	0.018	0.227**	1.333**
	(0.939)	(0.010)	(0.072)	(0.503)
turnover	1.697	0.044	−0.792	−7.008
	(6.948)	(0.087)	(0.460)	(6.051)
Overall R^2	0.230	0.285	0.320	
Within R^2	0.054	0.151	0.180	
F-Statistic	12.005***	17.994***	9.951***	
Log Likelihood				−350.233
AIC				710.466
BIC				744.402

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Clustered standard errors are in parentheses. The Amihud illituidity ratio is multiplied by 1000.

we estimate regressions with squared terms for all independent variables and present the results in Table A5 in the Appendix. Based on individual and joint significance, we then construct the regressions presented by Table 6.6.

The linear and quadratic terms for Garman-Klass volatility are jointly insignificant in three models. Only in the regression with average rank as the dependent variable is the linear term individually significant. The linear and quadratic components of the turnover ratio are also jointly insignificant for three models. They achieve joint significance only in the fixed effects logit. On the other hand, the logarithm of volume is significant across all four regressions. It negatively affects the average rank and positively impacts the average p-value as well as the probability of obtaining at least four tests that suggest a lack of predictability. Hence, there is more evidence in favor of a negative relationship between cryptoasset volume and return predictability. Nevertheless, when considering the model with the total number of tests as the dependent variable, a quadratic relation between volume and predictability is found. Initially, increases in the logarithm of volume lead to increases in the total number of tests. After the turning point is passed, the total number of tests that indicate a lack of predictability decreases as the logarithm of volume increases. Furthermore, similarly to other panel data regressions, one may observe that the within R^2 remains quite low.

Table 6.6: Panel Data Models (Subset)

		Linear FE		Logit FE
	rank	average p-value	total tests	
gk	239.721*			
	(98.818)			
gk²	−81.151			
	(42.910)			
ami	11.226*	−0.476*	−1.120	−26.986
	(5.085)	(0.234)	(2.602)	(20.044)
ami²		0.361	4.376**	13.681*
		(0.190)	(1.455)	(5.693)
spread				80.084*
				(37.198)
spread²				52.161
				(27.296)
logvol	−1.717*	0.023**	35.403**	153.362*
	(0.816)	(0.008)	(11.930)	(65.494)
logvol²			−20.790***	62.680
			(6.309)	(47.422)
turnover				−20.776
				(49.452)
turnover²				−34.605
				(26.511)
Overall R²	0.233	0.282	0.321	
Within R²	0.058	0.147	0.182	
F-Statistic	26.524***	21.731***	13.018***	
Log Likelihood				−326.189
AIC				668.377
BIC				722.675

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Clustered standard errors are in parentheses. The Amihud illituidity ratio is multiplied by 1000.

The linear terms of the Amihud illiquidity ratio are individually significant for models with the average rank and average p-values as dependent variables. The signs of the respective coefficients indicate a negative relationship between liquidity and return predictability. The regression with the total number of tests as the dependent variable as well as the fixed effects logit document a quadratic relation between the Amihud illiquidity ratio and the predictability measures. For these models, however, the coefficient of the squared term is positive. Additionally, the linear term of the Corwin-Schultz spread estimate is individually significant and positive in the fixed effects logit. These findings then lend only some support to the results of the cross-sectional analysis, which documents a negative relationship between cryptoasset liquidity and return predictability. Overall results from panel data models provide rather mixed evidence for the effect of liquidity on return predictability. Instead, the panel data analysis highlights the relationship between return predictability and cryptoasset volume.

6.3 Further Discussion and Limitations

We obtain mixed evidence about the effect of liquidity on cryptoasset predictability. Results from regressions on cross-sections suggest a negative impact of liquidity, measured namely by the Corwin-Schultz spread estimate. Results from panel data regressions, which include only the top 50 cryptoassets with the largest market capitalization, provide some support for a negative effect of liquidity, measured by the Amihud illiquidity ratio, on cryptoasset predictability. Yet when considering the full data set, results from panel data indicate no significant, negative relationship. Some of the regressions even present coefficients that suggest an opposite relation. Panel data, however, documents a negative effect of cryptoasset volume on the degree of return predictability. We, furthermore, find only limited support for a quadratic relationship between liquidity and predictability. Two subset panel data models and only one full panel data model provide evidence in favor of this relation. Altogether, we discover several main findings.

Firstly, a negative relationship between cryptoasset liquidity and return predictability is mainly supported by evidence from cross-sectional data. With this regard, findings from the cross-sectional analysis support the results of Wei (2018) as well as Brauneis and Mestel (2018). All cross-sectional regression specifications consistently document a negative impact of cryptoasset liquidity on the degree of return predictability. Panel data models, using only data from 50 cryptoassets with the largest market capitalization, provide some evidence in favor of this relation, as seen from Table 6.6. Yet, the fixed effects logit computed on the subset data estimates a contradictory coefficient for the Corwin-Schultz spread. Full panel data models do not provide evidence for a negative effect of liquidity on return predictability. Some specifications even document a significantly positive effect. Moreover, we do not find much evidence for a quadratic relationship between cryptoasset liquidity and return predictability. This relation is only supported by the fixed effects logit models and by the model with the total number of tests as the dependent variable, regressed on subset panel data. These models are visible in Tables 6.4 and 6.6. Altogether, we find mixed evidence about the relationship between cryptoasset liquidity and return predictability. With this respect, we discover discrepancies between the results from cross-sections and findings from panel data.

Secondly, for the cross-sectional models, the Corwin-Schultz spread estimate is mainly significant. This is in line with the results by Brauneis and Mestel (2018), who also document a significant relationship for the Corwin-Schultz spread, albeit the Amihud illiquidity ratio remains insignificant. Interestingly, when considering the panel data subset regressions, the evidence that is obtained relies mainly on the Amihud illiquidity ratio. This may thus be a further potential difference between cross-sectional and panel data results.

Thirdly, full panel data regressions provide evidence of a significant, negative relationship between the logarithm of volume and the degree of cryptoasset predictability. The relation is, moreover, supported by regressions with and without quadratic terms and holds for the full panel as well as for the subset. This is an interesting finding, indicating that cryptoassets, which are traded more, are also less predictable. It also suggests that the less traded, more niche cryptoassets are more predictable. Nevertheless, with this regard it may be noteworthy to again mention that although CoinMarketCap (2020) reportedly excludes from its metrics markets, which exhibit characteristics of artificially inflating traded volumes, potential concerns with the precision of reported figures should be noted. Furthermore, volume not only affects the *logvol* variable but also influences the Amihud illiquidity ratio, which uses traded volume in its calculation. It is thus interesting that both the logarithm of volume and the Amihud illiquidity ratio are important variables for panel data regressions. One may also note the contrast with the cross-sectional analysis, which does not give much significance to these two regressors and rather identifies the Corwin-Schultz spread and the turnover ratio as important independent variables.

Additionally, there is evidence of a quadratic relationship between Garman-Klass volatility and cryptoasset predictability for full panel data as well as evidence of a quadratic relationship between the turnover ratio and cryptoasset predictability for cross-sectional data. These results are limited to the respective data types, meaning that there is no quadratic effect of Garman-Klass volatility for cross-sectional data and no quadratic effect of the turnover ratio for panel data. This is thus a further discrepancy between the results for cross-section and panel data regressions.

As explained in Section 3.1, we refrain from directly measuring the impact of liquidity on market efficiency, instead concentrating on return predictability alone. In order to fully establish market efficiency or inefficiency an analysis of return predictability may not be enough. When investigating efficiency concerns further caveats prevail. For a comprehensive analysis of efficiency one should also take into account the feasibility of potential profit-generating trading strategies. For our analysis, we do not aim to examine risks of specific crypto-exchanges or any potential deposit, trading or withdrawal costs. Additionally, it may not be possible for some of the more obscure, less traded cryptoassets to be readily exchanged into US dollars or other fiat currencies. For some it may be necessary to first be exchanged into Bitcoin or another cryptoasset before being exchanged into a fiat currency. Once these costs are accounted for, return predictability may not be enough for a trading strategy to successfully obtain abnormal profit and thus pose as a violation of market efficiency. Hence, our findings are limited to cryptoasset return predictability and should not be generalized for market efficiency. We advise potential investigations into market efficiency for further related studies.

Future investigations may also analyze any potential differences between coins and tokens. One can examine high-frequency data, concentrate on different time frames or include different combinations of cryptoassets. Different measures for cryptoasset liquidity and return predictability may also be calculated. In our analysis, we include a range of predictability measures to ensure that our results would not depend on one measure only. Yet, our findings should be viewed in terms of the used predictability tests. Furthermore, an alternative approach may involve estimating concrete models. For instance, one can estimate autoregressive models and quantify their performance for different cryptoassets by the mean squared error. A thorough discussion and evaluation, explaining which models would be chosen, should be included. Moreover, as an exciting topic for future studies, one can predict cryptoasset returns using various machine learning methods. For example, Olson and Mossman (2003) report that back propagation neural networks provide superior forecasts of stock returns when compared with OLS and logit models. Neural networks have also been used to predict stock returns by Ticknor (2013) as well as Liao and Wang (2010). Interestingly, McNally, Roche, and Caton (2018) predict Bitcoin prices using two deep learning models and document their superior forecasting performance in comparison with an ARIMA specification. Cryptoasset predictability can thus be measured in different manners. Subsequently, a similar framework as the one presented in this thesis may be used to detect the relationship between cryptoasset liquidity and return predictability.

Chapter 7

Conclusion

Cryptoassets may provide an engaging and appealing investment opportunity. The intricacies of the crypto-markets continue to interest potential investors as well as researchers. Studies, examining the relationship between liquidity and return predictability within the crypto-markets, may then be enlightening. Past literature predominantly concentrates on the predictability of cryptoasset returns. Only a few studies document the relationship between cryptoasset liquidity and return predictability. Moreover, to the best of our knowledge, at the time of writing no previous study includes a panel data analysis that examines this relation. We thus investigate the relationship between cryptoasset liquidity and return predictability using both cross-sectional as well as panel data.

We collect daily market data for 100 cryptoassets during the years 2017 till 2019, thus obtaining a relatively rich data set with both widely traded and more niche cryptoassets for a time period that encompasses the boom on crypto-markets as well as the subsequent decline. To measure predictability we follow previous literature, namely Brauneis and Mestel (2018) as well as Wei (2018), and calculate a variety of predictability measures. Specifically, we consider the Bartels test, runs test, Ljung-Box test, difference sign test, turning point test, two Lo-MacKinlay variance ratio tests with different holding periods as well as the GPH estimator. Using these measures, we construct four dependent variables. One dependent variable is average cryptoasset rank, calculated based on the p-values from the predictability tests and the absolute value of the GPH estimate. Notably, all of the tests possess a null hypothesis that indicates a lack of predictability. The second dependent variable is the average p-value for each cryptoasset across the seven predictability tests. The third dependent variable consists of the total number of tests that suggest a lack of predictability. The fourth dependent variable is binary, equal to one when there are at least four such tests for a cryptoasset.

To gauge liquidity we include two measures, specifically the Amihud illiquidity ratio and the Corwin-Schultz spread estimate. The Amihud illiquidity ratio is a commonly used liquidity measure, while the Corwin-Schultz spread estimate allows us to gauge liquidity without having to rely on reported cryptoasset volume. Considering the uncertainty with the accuracy of reported cryptoasset volume figures and concerns that reported volumes for cryptoassets may be inflated (Singer, 2020), the spread estimate thus provides a particularly befitting method to account for liquidity. Other considered variables include the logarithm of volume, turnover ratio and Garman-Klass volatility.

Building on the study by Brauneis and Mestel (2018), we first examine the relationship between liquidity and predictability for cross-sectional data. We then add to existing literature and investigate the relation also using panel data. Furthermore, we examine the full panel data set as well as a restricted panel subset, which contains 50 cryptoassets with the largest market capitalization. We aim to investigate whether the relationship between liquidity and return predictability in the subset is similar to the relation in the full panel. For the cross-sectional analysis, we estimate OLS and logit models, while the panel data analysis includes linear fixed effects as well as logistic fixed effects models.

Results from the cross-sectional analysis provide evidence for a negative relationship between cryptoasset liquidity and return predictability. These findings mainly rely on liquidity as measured by the Corwin-Schultz spread estimate and support conclusions reached by Brauneis and Mestel (2018) as well as Wei (2018). The results from panel data, however, differ from the cross-sectional findings in numerous ways. For instance, when considering models calculated on the full panel, mixed evidence is obtained with respect to the impact of liquidity on cryptoasset predictability. Liquidity is either insignificant or possesses a positive effect on return predictability. When considering regressions computed on the subset of the panel data, however, there is some evidence in favor of a negative effect of liquidity on cryptoasset predictability, supporting the cross-sectional results. Overall, we find mixed evidence about the relationship between cryptoasset liquidity and return predictability.

Additionally, panel data regressions on both the full panel as well as the subset indicate a negative relationship between cryptoasset volume and return predictability, suggesting that the more traded cryptoassets are also less predictable. This finding may be interesting, considering the context of potential cryptoasset volume inflation. Moreover, there is evidence of a quadratic relationship between the turnover ratio and cryptoasset return predictability for cross-sections as well as evidence for a quadratic relationship between Garman-Klass volatility and predictability for the full panel. Both aspects could be subjects for future research.

Future studies may concentrate on different combinations of cryptoassets or different time periods. Potential differences between coins and tokens may also be analyzed. Moreover, different methods of predicting cryptoasset returns may be considered. For our analysis, we do not aim to investigate market efficiency concerns. Examining return predictability alone may not be enough to claim assertions about market efficiency. We thus recommend market efficiency concerns as a natural extension for further studies.

References

Alvarez-Ramirez, J., Rodriguez, E., & Ibarra-Valdez, C. (2018). Long-range correlations and asymmetry in the Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, *492*, 948–955.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Avramov, D., Chordia, T., & Goyal, A. (2006). Liquidity and Autocorrelations in Individual Stock Returns. *The Journal of Finance*, *61*(5), 2365–2394.

Badev, A., & Chen, M. (2014). *Bitcoin: Technical Background and Data Analysis* (Finance and Economics Discussion Series No. 2014-104). Washington, D.C.: Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board.

Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters*, *161*, 1–4.

Bartels, R. (1982). The Rank Version of von Neumann's Ratio Test for Randomness. *Journal of the American Statistical Association*, *77*(377), 40–46.

Brauneis, A., & Mestel, R. (2018). Price discovery of cryptocurrencies: Bitcoin and beyond. *Economics Letters*, *165*, 58–61.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (3rd ed.). Springer.

Broock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, *15*(3), 197–235.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets.* Princeton University Press.

Caporale, G. M., Gil-Alana, L., & Plastun, A. (2018). Persistence in the cryptocurrency market. *Research in International Business and Finance*, *46*, 141–148.

Chordia, T., Roll, R., & Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, *87*(2), 249–268.

CoinMarketCap. (2020). *Metric Methodologies.* Retrieved from `https://support.coinmarketcap.com/hc/en-us/sections/360008888252-Metric-Methodologies`

Corwin, S. A., & Schultz, P. (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance*, *67*(2), 719–759.

Cryptoassets Task Force. (2018). *Cryptoassets Task Force: final report* (Tech. Rep.). London: HM Treasury, the Financial Conduct Authority and the Bank of England.

Escanciano, J. C., & Lobato, I. N. (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics*, *151*(2), 140–149.

European Central Bank. (2019). *Cryptoassets – trends and implications.* Retrieved from `https://www.ecb.europa.eu/paym/intro/mip-online/2019/html/1906_crypto_assets.en.html`

European Central Bank Cryptoassets Task Force. (2019). *Cryptoassets: Implications for financial stability, monetary policy, and payments and market infrastructures* (European Central Bank Occasional Paper Series No. 223). Frankfurt am Main: European Central Bank.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Foley, S., Karlsen, J. R., & Putniņš, T. (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? *Review of Financial Studies*, *32*(5), 1798–1853.

Garman, M. B., & Klass, M. J. (1980). On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, *53*(1), 67–78.

Geweke, J., & Porter-Hudak, S. (1983). The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis*, *4*(4), 221–238.

Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric Statistical Inference* (4th ed.). Marcel Dekker.

Goyenko, R. Y., Holden, C. W., & Trzcinka, C. A. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, *92*(2), 153–181.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Pearson.

Hileman, G., & Rauchs, M. (2017). *Global Cryptocurrency Benchmarking Study* (Tech. Rep.). Cambridge: Cambridge Center for Alternative Finance, University of Cambridge.

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, *6*(2), 95–101.

Krištoufek, L. (2018). On Bitcoin markets (in)efficiency and its evolution. *Physica A: Statistical Mechanics and its Applications*, *503*, 257–262.

Krištoufek, L., & Vošvrda, M. (2014). Measuring capital market efficiency: long-term memory, fractal dimension and approximate entropy. *The European Physical Journal B: Condensed Matter and Complex Systems*, *87*(7), 1-9.

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

Köchling, G., Müller, J., & Posch, P. N. (2019). Price delay and market frictions in cryptocurrency markets. *Economics Letters*, *174*, 39–41.

Liao, Z., & Wang, J. (2010). Forecasting model of global stock index by stochastic time effective neural network. *Expert Systems with Applications*, *37*(1), 834–841.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Lo, A. W., & MacKinlay, A. C. (1987). *Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test* (NBER Working Paper Series No. 2168). Cambridge, MA: National Bureau of Economic Research.

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

McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)* (pp. 339–343). Institute of Electrical and Electronics Engineers.

Mersch, Y. (2018). *Virtual or virtueless? The evolution of money in the digital age* [Lecture]. Official Monetary and Financial Institutions Forum, London. (8 February)

Moore, G. H., & Wallis, W. A. (1943). Time Series Significance Tests Based on Signs of Differences. *Journal of the American Statistical Association*, *38*(222), 153–164.

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System.* Retrieved from `https://bitcoin.org`

Olson, D., & Mossman, C. (2003). Neural network forecasts of Canadian stock returns using accounting ratios. *International Journal of Forecasting*, *19*(3), 453–465.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, *6*(2), 41–49.

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

Singer, A. (2020). Cleaning Up Crypto Exchange Wash Trading Will Take Global Regulation. *Cointelegraph*. Retrieved from `https://cointelegraph.com/news/cleaning-up-crypto-exchange-wash-trading-will-take-global-regulation`

Stammann, A., Heiss, F., & McFadden, D. (2016). *Estimating Fixed Effects Logit Models with Large Panel Data* [Annual Conference 2016 (Augsburg): Demographic Change]. Verein für Socialpolitik / German Economic Association.

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications*, *40*(14), 5501–5506.

Timmermann, A., & Granger, C. W. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting*, *20*(1), 15–27.

Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, *148*, 80–82.

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

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). South-Western Cengage Learning.

Yermack, D. (2013). *Is Bitcoin a Real Currency? An economic appraisal* (NBER Working Paper Series No. 19747). Cambridge, MA: National Bureau of Economic Research.

Zhang, W., Wang, P., Li, X., & Shen, D. (2018). The inefficiency of cryptocurrency and its cross-correlation with Dow Jones Industrial Average. *Physica A: Statistical Mechanics and its Applications*, *510*, 658–670.

Appendix

Table A1: List of Cryptoassets

btc	strat	block	pot	mint
eth	gnt	nav	zcl	xst
xrp	maid	nmc	flo	cure
ltc	dgb	grs	xpm	thc
dash	ardr	nlg	pasc	glc
xmr	fct	xcp	grc	sphr
xlm	xvg	emc2	rads	obits
etc	pivx	burst	xwc	pnd
xem	mona	via	crw	trc
neo	nxt	xdn	dmd	cann
zec	gbyte	sls	vrc	ixc
doge	sys	uno	omni	egc
lsk	xzc	edc	sib	orb
waves	rdd	lbc	slr	cpc
dcr	nxs	aeon	mue	xmg
rep	vtc	clam	exp	zet
bcn	emc	blk	ok	xcn
steem	game	cloak	nvc	dgc
bts	ppc	ioc	bcy	vrs
sc	sngls	ftc	xmy	swing

Note: The cryptoassets are listed from largest to smallest in terms of market capitalization.

Table A2: Average Partial Effects for Logit Models

	Table 6.1	Table A3	Table 6.2	Table 6.3	Table 6.5
gk	0.206	0.149		0.046***	−0.011
	(0.145)	(0.218)		(0.013)	(0.031)
ami	−3.863	10.365		0.013	0.006
	(10.734)	(44.862)		(0.017)	(0.065)
spread	−2.868***	−1.581	−2.388***	0.537***	0.091
	(0.811)	(1.325)	(0.323)	(0.130)	(0.238)
logvol	0.043	0.029		0.020***	0.021
	(0.024)	(0.033)		(0.005)	(0.014)
turnover	−0.098	8.321**	7.227***	−0.029	−0.110
	(1.028)	(2.879)	(1.560)	(0.109)	(0.175)

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Standard errors are in parentheses. Clustered standard errors are used for panel data models (Tables 6.3 and 6.5). The Amihud illiquidity ratio in Table 6.5 is multiplied by 1000.

Table A3: Cross-Section Models with Quadratic Terms

	OLS			Logit
	rank	average p-value	total tests	
Constant	50.728***	0.233***	3.980***	−125.274
	(1.241)	(0.010)	(0.143)	(544.424)
gk	−17.104	0.033	4.344	−105.931
	(39.587)	(0.306)	(4.549)	(64.742)
gk^2	21.211	0.015	−3.343	−89.612
	(24.668)	(0.174)	(2.835)	(54.693)
ami	5.171	0.032	0.027	−10,216.880
	(15.695)	(0.045)	(1.804)	(45,754.620)
ami^2	−6.095	0.020	0.481	−2,839.108
	(13.887)	(0.035)	(1.596)	(12,671.520)
spread	104.898***	−0.513*	−14.589***	−42.036*
	(30.577)	(0.224)	(3.514)	(20.913)
$spread^2$	−7.070	−0.025	2.649	−15.715
	(18.744)	(0.111)	(2.154)	(18.142)
logvol	−12.358	0.319	1.818	11.384
	(32.119)	(0.242)	(3.691)	(12.167)
$logvol^2$	4.857	0.110	−1.701	6.472
	(15.207)	(0.144)	(1.748)	(8.709)
turnover	−25.725	0.077	3.032	10.601
	(17.264)	(0.119)	(1.984)	(7.598)
$turnover^2$	40.746**	−0.420***	−5.559***	−16.729**
	(14.175)	(0.092)	(1.629)	(6.025)
R^2	0.506	0.434	0.549	
Adjusted R^2	0.450	0.370	0.499	
F Statistic	9.105***	6.813***	10.847***	
Log Likelihood				−33.094
AIC				88.187
BIC				116.844

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Standard errors are in parentheses.

Table A4: Panel Data Models with Quadratic Terms

	Linear FE			Logit FE
	rank	average p-value	total tests	
gk	−37.781	−0.154	41.993	123.992**
	(252.275)	(3.068)	(25.187)	(42.796)
gk^2	−217.855**	1.990**	18.463**	25.320
	(70.508)	(0.765)	(6.723)	(15.970)
ami	−45.552	0.839*	8.852	−6.932
	(27.601)	(0.350)	(4.682)	(11.507)
ami^2	2.857	−0.120	1.672	18.191*
	(22.687)	(0.314)	(4.414)	(8.217)
spread	−410.998	3.275	83.927***	187.216***
	(238.442)	(2.922)	(25.313)	(51.032)
$spread^2$	−84.661	1.201	−1.149	0.490
	(88.883)	(0.919)	(8.254)	(14.352)
logvol	−815.752**	8.999**	94.696***	209.554***
	(259.598)	(3.062)	(26.289)	(52.774)
$logvol^2$	−66.491	1.375	−0.385	91.594**
	(145.774)	(1.624)	(14.494)	(28.236)
turnover	23.817	−0.185	−6.074	−81.470
	(93.623)	(1.126)	(8.418)	(63.969)
$turnover^2$	34.524	−0.359	−4.360	−76.612
	(49.419)	(0.560)	(4.506)	(68.304)
Overall R^2	0.415	0.415	0.525	
Within R^2	0.060	0.150	0.203	
F-Statistic	4.700***	7.701***	8.714***	
Log Likelihood				−1733.394
AIC				3486.788
BIC				3561.592

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Clustered standard errors are in parentheses.

Table A5: Panel Data Models with Quadratic Terms (Subset)

	Linear FE			Logit FE
	rank	average p-value	total tests	
gk	252.527	−2.133	−6.595	−9.513
	(183.644)	(2.199)	(11.613)	(41.366)
gk^2	−86.020*	0.354	0.821	5.716
	(40.740)	(0.543)	(3.162)	(17.612)
ami	31.629	−0.312	−1.754	−26.544
	(31.989)	(0.344)	(2.695)	(22.278)
ami^2	−24.219	0.223	3.297*	13.996*
	(18.088)	(0.223)	(1.364)	(5.623)
spread	49.378	−0.332	9.403	71.111
	(217.174)	(2.417)	(13.990)	(68.594)
$spread^2$	22.665	−0.411	−1.061	52.227
	(38.483)	(0.424)	(6.171)	(26.844)
logvol	−349.300	3.102	34.050*	151.406*
	(215.035)	(2.553)	(16.133)	(65.742)
$logvol^2$	68.135	−0.597	−14.752	61.907
	(106.933)	(1.243)	(7.764)	(46.976)
turnover	4.293	0.605	−0.676	−21.296
	(78.521)	(0.950)	(5.089)	(49.821)
$turnover^2$	0.973	−0.206	−3.080	−35.085
	(43.717)	(0.503)	(2.866)	(26.091)
Overall R^2	0.234	0.287	0.327	
Within R^2	0.060	0.153	0.189	
F-Statistic	13.892***	10.167***	7.022***	
Log Likelihood				−325.941
AIC				671.882
BIC				739.755

Note: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Clustered standard errors are in parentheses. The Amihud illituidity ratio is multiplied by 1000.