FUNCTIONAL ANALYSIS OF INTRADAY AND DAILY DYNAMICS OF CRYPTOCURRENCY

Cheng Zhong

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
GRADUATE PROGRAM IN ECONOMICS
YORK UNIVERSITY
TORONTO, ONTARIO

May 2026

Abstract

Cryptocurrencies have become one of the fastest-growing asset classes with significant investment potential. This research examines the dynamics of cryptocurrency returns and prices, and introduces new methods of functional time series analysis. It reveals the existence of common periodic effects, which are recurrent intraday and intraweek patterns in hourly and daily prices, returns, volumes, and volatility of different types of cryptocurrency. The empirical results show that the common intraday periodicity in native cryptocurrencies and tokens is driven by the operating hours of the NYSE, LSE, and Han Seng stock exchange markets. Moreover, the intraday and daily cross-sectional correlation matrices of cryptocurrency returns indicate high correlation among native cryptocurrencies and tokens, while stablecoins exhibit distinct dynamics and are uncorrelated with other cryptocurrencies. A functional CAPM is introduced to accommodate these periodic patterns, and it is estimated by regressing the functions of intraday and intraweek cryptocurrency returns on the market portfolio. A PCA-based principal eigenportfolio of the correlation matrix is used as a proxy of the cryptocurrency market portfolio. The major native cryptocurrencies and tokens satisfy affine relationships with the return functions of the market portfolio, and their functional betas display periodic intraday and intraweek patterns.

To bridge the functional and discrete-time approaches, the Karhunen-Loeve (KL) Dynamic Factor model is introduced. It offers a convenient framework for analyzing conditionally heteroscedastic functional time series. Improved interval forecasts of 15-minute and hourly Bitcoin future return functions are obtained when the conditional heteroscedasticity of return functions is accounted for. A "Rolling" FPCA-based algorithm is introduced for forecasting intraday incomplete functions. The proposed algorithm improves the intraday hourly forecast of Bitcoin returns.

Acknowledgments

I would like to express my deepest gratitude to my PhD supervisor, Professor Joann Jasiak, for her mentorship, guidance, support, and encouragement throughout my doctoral studies at York University. Her continuous support and advice have been invaluable throughout my study journey.

I am also grateful to my fellow PhD colleagues at York University for their friendship, encouragement, and support during my studies.

Finally, I would like to thank my parents, my wife, and my son for their unconditional love, patience, and support throughout this journey. Their love and encouragement have given me strength and motivation throughout my doctoral journey.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cryptocurrencies have become among the most important and rapidly expanding assets since Bitcoin began trading in 2009. Since November 2024, the total cryptocurrency market cap has exceeded 3 trillion (TradingView, 2026). With the rapid development of computational power and the availability of high-frequency data, the intraday and daily dynamics of cryptocurrencies have attracted considerable attention from researchers and investors.

The motivation for this thesis is to improve the understanding of the intraday and intraweek dynamics and predictability of cryptocurrencies by using the functional time series methodology. The empirical results presented in this thesis reveal common patterns in the prices, returns, volumes, and volatility of cryptocurrencies, such as native cryptocurrencies, stablecoins, and tokens. The exposure of individual cryptocurrencies to systemic risk is examined by modeling their returns as an affine function of returns on the cryptocurrency market portfolio. The thesis makes additional methodological contributions by improving interval forecasting of complete intraday return functions and point forecasting of incomplete intraday return functions of cryptocurrencies.

This thesis comprises three closely connected chapters, each of which examines one of the above research topics. The three chapters together provide a coherent framework for analyzing the intraday and daily dynamics and predictability of the cryptocurrency market.

Chapter 2 analyzes the intraday and daily dynamics of cryptocurrencies of different types. Common intraday and intraweek patterns in prices, returns, volumes, and volatility are documented among native cryptocurrencies and tokens. The opening and closing times of major stock exchanges play a significant role in shaping intraday patterns. In contrast, comparable systematic patterns are not observed in stablecoins. In Chapter 3, the functional Capital Asset Pricing Model for cryptocurrency is introduced, using a PCA-based eigenportfolio as a proxy for the market portfolio. This approach offers a comprehensive analysis of the relationship between individual cryptocurrency functional returns and cryptocurrency market portfolio returns.

The fourth chapter contributes methodologically to the econometric literature by introducing a Karhunen-Loeve (KL) Dynamic Factor model that bridges the conditionally heteroscedastic discrete-time models with their functional counterparts. The proposed model delivers improved pointwise interval forecasts for intraday Bitcoin return functions. Secondly, a rolling FPCA algorithm is introduced to enable incomplete intraday point forecasting of Bitcoin returns.

Together, these three chapters make substantive empirical and methodological contributions to the literature. They provide empirical evidence of common trends among native cryptocurrencies and tokens, while such patterns are absent in stablecoins. The documented relationship between individual cryptocurrency returns and returns on the cryptocurrency market portfolio offers convincing evidence of systemic market risk. Moreover, the KL dynamic factor model provides a new functional framework for improved interval forecasting of return functions.

The remaining part of the thesis is structured as follows: Chapter 2 investigates the intraday and daily trading behavior across different types of cryptocurrency. Chapter 3 introduces the functional CAPM and examines the relationship between individual cryptocurrency returns and the returns on the cryptocurrency market portfolio. Chapter 4 proposes the KL dynamic factor models and the rolling FPCA model for forecasting Bitcoin return functions.

Chapter 5 concludes by summarizing key findings, empirical results, and directions for future research.

Chapter 2

Intraday and Daily Trading Behavior in Cryptocurrency Markets

Co-authorship Statement

This chapter is based on joint work titled "Intraday and Daily Dynamics of Cryptocurrency" with Dr. Joann Jasiak. All authors contributed equally to the conceptualization, data preparation, empirical modeling, analysis, and writing of the manuscript.

A version of this chapter is currently accepted and published at `International Review of Economics and Finance`.

2.1 Introduction

Many financial assets exhibit intraday and intraweek periodicity characterized by recurrent patterns in their prices, returns, volumes, and volatility. Among them are digital assets, i.e. cryptocurrencies. In this class, we distinguish native cryptocurrencies that have their own blockchains, such as Bitcoin and Ethereum, tokens, which are built on already existing blockchains, such as LINK, and stablecoins, whose values are fixed to another asset, such as Tether pegged to the US Dollar (USD). Since 2021, the cryptocurrency market has evolved towards becoming more transparent, better organized, and regulated, while its capitalization surpassed 1 Trillion USD. It also became more heterogeneous, as the part of market capitalization of stablecoins and tokens has increased. While Bitcoin and Ether are remaining the top two coins with the highest capitalizations, the stablecoin Tether (USDT) moved to third place. Currently, there are two stablecoins and five tokens among the top ten coins. Various cryptocurrency indexes have been created for tracking their performance, such as the S&P Cryptocurrency MegaCap Index of cryptocurrencies with the highest capitalization and the Standard and Poor's Cryptocurrency Broad Digital Market (SPCBDM) index for digital assets listed on recognized open digital exchanges. These developments mark the beginning of a new episode in cryptocurrency markets, which motivates our study aimed at describing and comparing the periodicity in the dynamics of native cryptocurrency, stablecoins and tokens.

The knowledge of periodic patterns, either common or specific to each type of cryptocurrency, is important to investors for portfolio diversification and for designing effective trading strategies adapted to the native cryptocurrencies, tokens and stablecoins. Identifying the recurrent dynamic patterns can offer researchers a better understanding of the market behavior and facilitate efficient asset pricing and systemic risk assessment by financial market analysts. Moreover, by exploiting the predictable patterns, it is possible to develop or improve the performance of dynamic models and forecasts.

Our first contribution is in documenting the common intraday and intraweek patterns in the prices, returns, volumes, and volatility of native cryptocurrency and tokens, and in emphasizing a distinct behavior of stablecoins. The novelty of our approach is in considering a diverse sample of cryptocurrencies frequently traded on Bitstamp, rather than Bitcoin only, to offer more comprehensive evidence of periodicity than the existing literature. We find that intraday periodic patterns are strongly influenced by the opening and closing times of major stock exchange markets: NYSE, LSE, and Hang Seng, confirming the existence of interactions between the cryptocurrency and stock markets that can be exploited for diversified investments. For example, investors can benefit from lower prices of natives and tokens between 4 pm and 8 pm Eastern Standard Time, i.e., between the closing time of NYSE and the opening time of Hang Seng. Hang Seng opening is associated with increased prices and positive intraday returns. In daily data, we find that prices and returns of native cryptocurrency and tokens are decreasing over the trading week to their lowest levels on Friday.

As mentioned earlier, our analysis is focused on the cryptocurrencies traded on Bitstamp (Bitstamp, 2022), including the natives, stablecoins and tokens in 2022. Our findings, in part, corroborate and extend the results of Eross et al. (2019) and Ma and Tanizaki (2019), who examined the periodicity in Bitcoin from intraday and intraweek data, respectively, recorded before the Covid-19 pandemic and prior to the associated expansion of the cryptocurrency market. More specifically, Eross et al. (2019) studied intraday 5-minute returns, liquidity, and volatility of Bitcoin data from www.bitcoincharts.com between January 2014 and December 2017. They documented periodic patterns in Bitcoin returns, volatility, and liquidity. They observed the impact of the openings of major North American and European currency markets on Bitcoin, which we also detected in the common trading patterns of natives and tokens. In addition, we evidence the effects of the Chinese market, which are shown to be significant in our study. Ma and Tanizaki (2019) studied daily data on Bitcoin from Coindesk between January 2013 and December 2018, and described the day-of-the-week effects in the returns and volatility. They documented higher volatility of Bitcoin on Mondays, which is in line with our results on Bitcoin. The periodicity in Bitcoin intraday and intraweek volatility and volume was examined in a more recent study of Hansen et al. (2024) based on intraday 1-minute data from two centralized exchanges (Coinbase Pro and Binance) and a decentralized exchange (Uniswap V2) over the period October 2020 to June 2021. This paper evidenced the day-of-the-week and hour-of-the-day effects in the volatility and volumes of Bitcoin and Ether without associating them with the operating times of major stock markets. Selected volatility and volume patterns of Bitcoin reported in Hansen et al. (2024) are close to those documented in our sample, such as lower Bitcoin volumes in the first part of UTC day and over the weekends. The calendar anomalies in Bitcoin volatility were also reported in Kinateder and Papavassiliou (2021). Earlier papers on Bitcoin by Fraz et al. (2019) and Caporale and Plastun (2019) considered daily data on Bitcoin returns from 2013 to 2017 and evidenced the Monday effect, i.e., higher Monday returns on Bitcoin, in line with Ma and Tanizaki (2019).

2.2 Data description

The data source is Bitstamp, the biggest European cryptocurrency exchange platform, which has been operating since 2011, and is recommended as a reliable data provider in Vidal-Tomás (2022). Our intraday data are sampled hourly over the period of 6 months, starting from 2022-01-01 until 2022-06-28, with a total number of 179 days[1]. The daily data are observed over one year, 2022-01-01 to 2022-12-30, with a total number of 52 weeks of data on the adjusted closing prices in US Dollars at 0:00 hour of the Coordinated Universal Time (UTC). The volumes of trades are defined in terms of the number of coins traded.

The UTC is the time scale of reference. For comparison, we provide below the local operating times of the major stock markets, which are as follows:
The US markets (NYSE, NASDAQ): 9:30 to 16:00 of Eastern Standard Time (EST). The UTC equivalent of EST time is obtained by adding 4;
London Stock Exchange (LSE): 8:00 to 16:30 of Greenwich Mean Time (GMT). The GMT is equivalent to the UTC time;
Hong Kong market time (Hang Seng): 9:30 to 16:00 with a break from 12:00 to 13:00 of Hong Kong Time (HKT). The UTC equivalent of HKT time is obtained by subtracting 8;

There exist small differences due to the daylight saving time. Table 2.1 compares the 24-hour scale of UTC time with Eastern Standard Time (ETS), Hong Kong Time (HKT) and Greenwich Mean Time with daylight saving (panel A) and without it (panel B). We

[1]We study periodic patterns in the year 2022 associated mostly with winter time (panel B of Table 1).

observe that the opening hours of NYSE and LSE overlap by 3 hours. The LSE overlaps with Hang Seng by one hour and only when Eastern Daylight Saving Time is implemented.

Table 2.1: Intraday timeline: UTC clock and local time of major markets

UTC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Panel A																								
NYSE	-20	-21	- 22	-23	0	1	2	3	4	5	6	7	8	***9***	***10***	***11***	***12***	***13***	***14***	***15***	16	17	18	19
LSE	1	2	3	4	5	6	7	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	17	18	19	20	21	22	23	+0
Hang Seng	8	***9***	***10***	***11***	12	***13***	***14***	***15***	16	17	18	19	20	21	22	23	+0	+1	+2	+3	+4	+5	+6	+7
Panel B																								
NYSE	-19	-20	-21	- 22	-23	0	1	2	3	4	5	6	7	8	***9***	***10***	***11***	***12***	***13***	***14***	***15***	16	17	18
LSE	0	1	2	3	4	5	6	7	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	17	18	19	20	21	22	23
Hang Seng	8	***9***	***10***	***11***	12	***13***	***14***	***15***	16	17	18	19	20	21	22	23	+0	+1	+2	+3	+4	+5	+6	+7

2.2.1 The Variables of Interest

The Bitstamp prices and volumes, as well as the series of returns and realized volatility, are used to compute the intraday and intraweek averages, following the approach of Eross et al. (2019).

2.2.1.1 Hourly Data

The hourly data are indexed by t, t = 0,1,..., 23, with t=0 at UTC hour 00:00 on the day d, d = 1, ..., 179, which is the number of days in the sample. The variables of interest are defined as follows:

1. The average intraday price over 179 days at time t: $\bar{P}_t = \frac{1}{179}\sum_{d=1}^{179} P_{t,d},\ t = 0, 1, ..., 23$

2. The intraday log-return: $r_{t,d} = (lnP_{t,d} - lnP_{t-1,d}) \times 100$, with the average over 179 days at hour t: $\bar{r}_t = \frac{1}{179}\sum_{d=1}^{179} r_{t,d},\ t = 0, 1, ..., 23$

3. The intraday realized volatility (RV) at hour t: Realized variance$_{t,d} = \frac{1}{179}\sum_{d=1}^{179}\left(ln\frac{P_{t,d}}{P_{t-1,d}}\right)^2 \times$ 10000. The square root of the realized variance at hour t is: $RV_t = \sqrt{\text{Realized variance}_{t,d}},\ t = 0, 1, ..., 23$

4. The average intraday volume at hour t: $\bar{v}_{t,d} = \frac{1}{179}\sum_{d=1}^{179} volume_{t,d}$

2.2.1.2 Daily Data

The daily variables are indexed by the days of the week denoted by n (n = 1,2,..,7 for Monday, Tuesday, ..., Sunday). The weekly averages are computed over 52 weeks of the year 2022. For days $n = 1, 2, ..., 7$ of the week, we define:

5. The average daily price over 52 weeks, $\bar{P}_n = \frac{1}{52}\sum_{j=1}^{52} P_{n,j},\ n = 1, .., 7$

6. The average daily return over 52 weeks, $\bar{r}_n = \frac{1}{52}\sum_{j=2}^{52} r_{n,j},\ n = 1, .., 7$ where $r_{n,j} = (lnP_{n,j} - lnP_{n,j-1}) \times 100$

7. The square root of RV over 52 weeks, $RV_n = \sqrt{\text{Realized variance}_{n,j}},\ n = 1, .., 7$ where Realized variance$_n = \frac{1}{52}\sum_{j=2}^{52}\left(ln\frac{P_{n,j}}{P_{n,j-1}}\right)^2 \times 10000$

8. The average daily volume over 52 weeks, $\bar{v}_n = \frac{1}{52}\sum_{j=1}^{52} volume_{n,j}$

Our data contains records on 45 cryptocurrencies, including 10 native coins, 6 stablecoins and 29 tokens [Sehra et al. (2018) and Grobys et al. (2021)]. The stablecoins considered in our study are pegged 1:1 to the US Dollar, except for EURT, which is pegged 1:1 to the Euro. For the empirical analysis, we consider an additional important characteristic of coins, which is their trading frequency and volume at the hourly and daily frequencies, reported in Tables A.1 and A.2, Appendix A. Based on these results, we distinguish the frequently traded coins at the hourly frequency (with more than 50% of non-zero volumes in the sample) and frequently traded coins at the daily frequency (with more than 90% of non-zero volumes in the sample) [2]. All natives are frequently traded at both the hourly and daily frequencies. These are the big-cap natives BTC, ETH and XRP with a market capitalization of over 10 billion Dollars and small-cap natives XLM, LTC, BCH, ADA, ALGO, SGB and HBAR.

At the hourly frequency, only the two big-cap stablecoins, Tether and USD Coin, with the largest capitalizations in the market, are frequently traded. At the daily frequency, three stablecoins: Tether, USD Coin and DAI trade sufficiently frequently.

Among the tokens, which are small-cap coins, only 18 satisfy the frequent trade requirement in hourly data, while 23 tokens are frequently traded at the daily frequency.

[2] In the case of zero volumes, the prices are considered constant.

We distinguish the Decentralized Finance (DeFi) Platforms and Tokens: Aave (AAVE), Compound (COMP), Curve Finance (CRV), Synthetix (SNX), Maker (MKR), Uniswap (UNI), Yearn.finance (YFI), UMA (UMA), Alpha Finance (ALPHA) and Perpetual Protocol (PERP); the Decentralized Oracle Network Chainlink (LINK); and the Blockchain-Based Gaming and NFTs: The Sandbox (SAND), Axie Infinity (AXS), Enjin Coin (ENJ), Smooth Love Potion (SLP) and Gala (GALA).

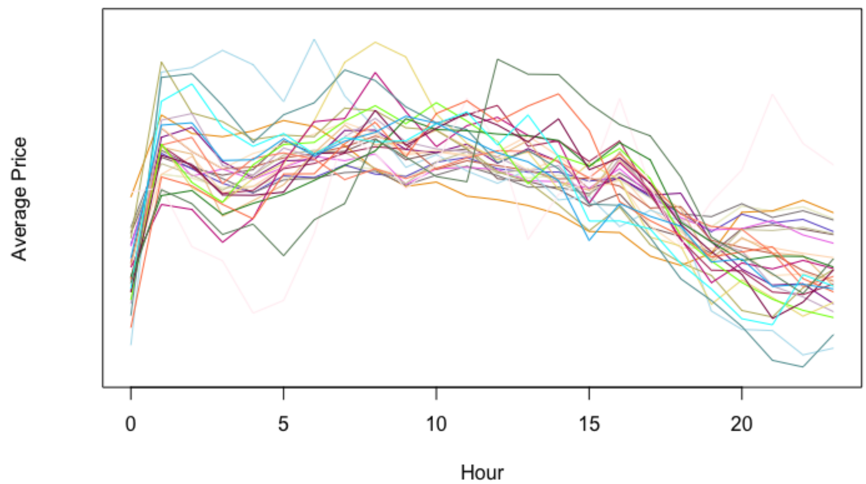
2.3 Intraday and Intraweek Patterns

This section describes and compares the common intraday and intraweek patterns of prices, returns, volumes and volatility of native cryptocurrency, stablecoins and tokens. In the first part, we examine the intraday data by computing the intraday averages defined in Section 2.1.1. The second part illustrates the results for intraweek data, based on the averages defined in Section 2.1.2. In each part, we describe and compare the common patterns of natives and tokens and report the behavior of stablecoins.

2.3.1 Intraday Patterns

The intraday dynamics of hourly prices of cryptocurrency motivating this research is illustrated in Figure 2.1, where the curves represent the prices of all frequently traded natives and tokens rescaled by their own daily price averages plotted over 24 hours of a UTC day.

Figure 2.1: Hourly Average (Rescaled) Prices of Frequently Traded Native Coins and Tokens (No Stablecoins)



All hourly average prices of coins are rescaled (but not demeaned) by their own averages, e.g., BTC's hourly average prices are rescaled (but not demeaned) by the average price of BTC between UTC hours 0 to 23.
We observe that the prices of natives and tokens increase at UTC hour 1:00 and remain high until UTC hour 14:00

We observe that the (rescaled) prices of frequently traded cryptocurrencies are relatively steady across the first part of UTC day and decrease after UTC hour 14:00 corresponding to the opening of NYSE [3]. The stablecoin prices are not reported in Figure 2.1 as they would appear constant because their volatility is much lower than the volatility of other coins, as evidenced later in Section 3.1.3.

2.3.1.1 Native cryptocurrency

By computing the hourly averages of prices, returns, volumes and volatilities, we find several common patterns in the intraday behaviour of the natives.

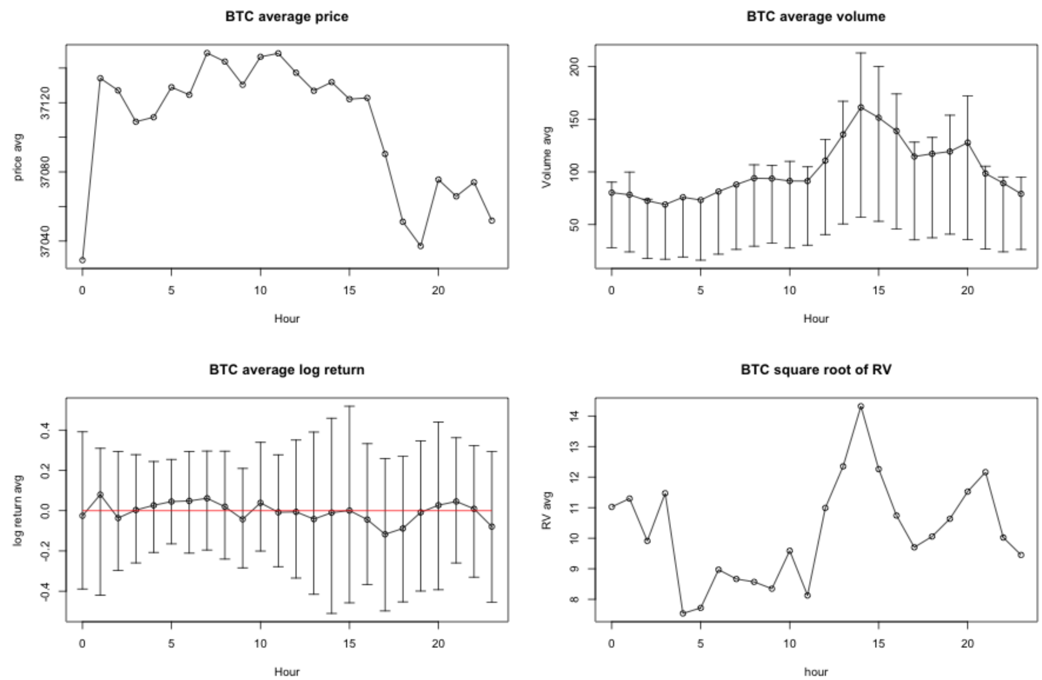
[3] We use panel B of Table 2.1 as the time of reference for intraday pattern description.

The hourly average prices of native cryptocurrencies tend to decrease between the NYSE opening of UTC hour 14:00 until the end of UTC day. This price decrease is associated with high volatility of returns at UTC hour 14:00, as indicated by a large intra-quartile range of returns, and peaks in both daily RV volatility and volume. The price rebounds slightly in some natives at UTC hour 20:00, which coincides with the NYSE closing time, except for ALGO, SGB, and HBAR. An increase in volatility and volume of natives is reported at UTC hour 20:00 as well. Afterwards, the prices, volumes and volatilities tend to remain low until the price increases sharply at the opening of Hang Seng at UTC hour 1:00. Another, although smaller price increase is observed at UTC hour 5:00 (except for SGB), which coincides with the re-opening of the second part of Hang Seng trading day.

Positive average returns tend to be associated with the price increases at UTC hours 1:00 (opening hour of Hang Seng), 5:00 (second part of Hang Seng day opening) (observed in Bitcoin (BTC), XLM, XRP, ETH, LTC, ADA, ALGO, HBAR) and 20:00 (NYSE closing hour) (observed in BTC, XLM, XRP, ETH, LTC, ADA). During the NYSE trading hours (UTC hours 14:00 to 20:00), the hourly average returns of native cryptocurrencies are mostly close to 0 or negative.

We find that most of the natives display their highest hourly average volumes during the overlap of the operating times of LSE and NYSE (UTC hours 14:00 to 16:00). The exceptions are XRP, SGB and HBAR. The hourly average volume of XRP peaks at UTC hour 10:00 and decreases after the LSE closing time. According to Bitstamp (https://www.bitstamp.net), XRP is unavailable in the US, which may explain this pattern. The hourly average volumes of SGB and HBAR peak at UTC hour 19:00, i.e. one hour before NYSE closes. The RV of most of the natives (except for SGB) is high during UTC hours 12:00 to 16:00, i.e., the second part of the LSE trading day. For most of the natives, this period of high volatilities is associated with high volumes and close to 0 returns.

Figure 2.2: BTC



The graph displays for each hour of the day: the average prices of Bitcoin (top left), the average volumes of Bitcoin (top right), the average returns of Bitcoin (bottom left) and the volatility of Bitcoin (bottom right). The vertical lines in the average volumes and log returns plots indicate the 25th and 75th percentiles of sample density. The zero value of average log returns is marked with a red line.

As an illustration, Figure 2.2 displays the average patterns in Bitcoin (BTC). The hourly prices of BTC are relatively lower after the closing hour of the NYSE, while volume and volatility are relatively higher during the overlapping hours of the LSE and NYSE. We confirm the pattern in volume that peaks around 14:00 GMT reported by Eross et al. (2019). However, the volatility and volume of Bitcoin in our sample tend to decrease later, i.e., after UTC hour 20 rather than after UTC hour 17. Compared with Hansen et al. (2024), high volatility of Bitcoin is observed earlier, i.e., around UTC hour 14 rather than around UTC hour 16. We also find that RV takes its lowest value around UTC hour 5 and we detect lower volumes of Bitcoin before UTC hour 12, which is in agreement with Hansen et al. (2024).

We find statistically significant differences between the returns on Bitcoin at UTC hour 1 (opening of Hang Seng), UTC hour 5 (opening of the second part of Hang Seng trading day), and UTC hour 20 (closing of NYSE). We use the Mehta and Gurland (1969) test to check if the means of hourly returns are statistically significantly different. For BTC, we find evidence against the null hypothesis: "the means are equal" when the BTC mean return at UTC hour 1 is compared with the mean return at UTC hours 4, 5, 7, 8, 10, 11, 20, and 22. The same conclusion holds for the mean return at UTC hour 5 compared with the mean return at hours 1, 17, 18, and 23 and the mean return at UTC hour 20 compared with hours 0, 1, 6, 12, 17, 18, 19, and 23. [4]

The Levene's test for homogeneity of return variance provides us with evidence against the null hypothesis of equal variances of returns at hour 1 compared with hours 4, 5, 6, 7, 8, 9, 10, 11, 14 15. We also find statistically different return variances at hour 5 compared with hours 0, 1, 2, 3, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23, as well as at hour 20 compared with hours 2, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 14.

The Kolmogorov-Smirnov test of return distributions provides us with evidence against the null hypothesis, "the samples were drawn from the same distribution" when we compare the returns at hour 1 with hours 3, 4, 5, 6, 7, 8, 10, 20 and 21. Statistically significant differences in return distributions are also found when the distributions of returns at hour 5 are compared with the return distributions at hours 0, 1, 4, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22 and 23. Furthermore, the return distributions at hour 20 are distinct from the return distributions at hours 1, 4, 5, 7, 9, 11, 12, 17, 18 and 23. [5]

2.3.1.2 Tokens

Common intraday patterns exist in the prices, returns, volumes, and volatility of tokens, which are documented below.

A common periodic pattern of tokens (except for KNC) is a price decrease following the opening of NYSE at hour 14:00 UTC, which also characterizes the prices of natives. The

[4] The critical level $\alpha = 0.1$.

[5] The tests (except for Mehta-Gruland) are applied under the assumption of uncorrelated returns. We find evidence of statistically significant correlations up to lag 10 in 4 out of 24 hours of BTC returns.

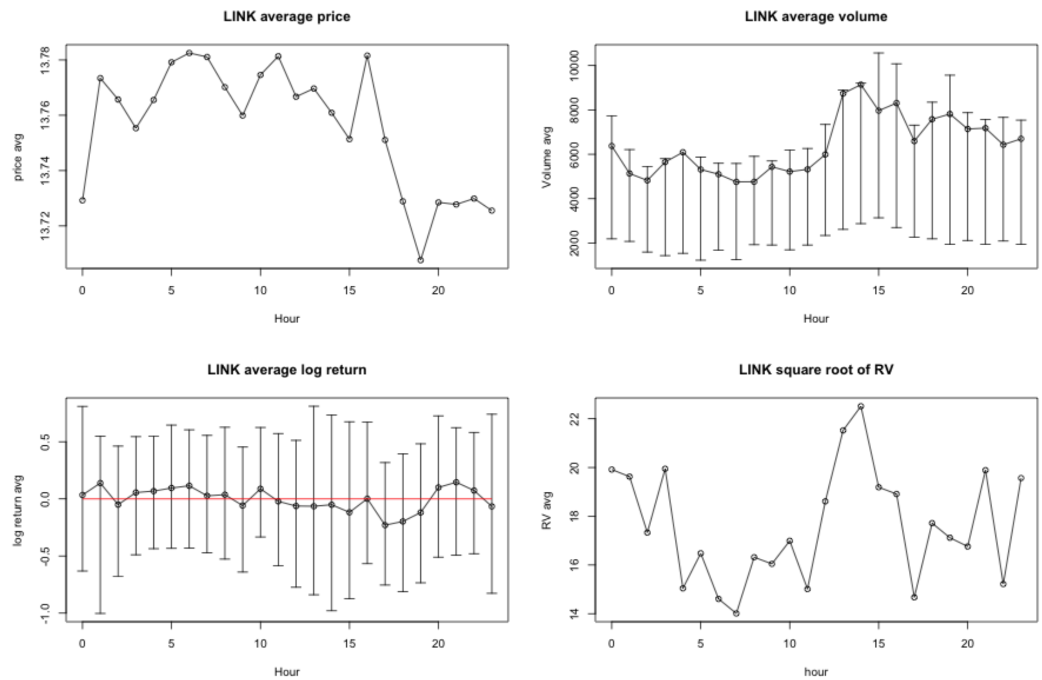
prices tend to be low after the NYSE closes at UTC hour 20:00 until the opening of Hang Seng at UTC hour 1:00 when they increase sharply. The prices of most of the tokens, like the natives, increase again at the re-opening of the second part of the Hang Seng trading day at UTC hour 5:00, which leads to positive average returns in most of the tokens [6]. There is another slight rebound observed in the prices of several tokens (e.g., LINK, UNI, SAND, AAVE, CRV, SNX, YFI, GRT, MKR) at UTC hour 16:00, which corresponds to the closing of LSE. Similarly to the natives, some of the Decentralized Finance (DeFi) and Financial Services tokens, such as AAVE, AUDIO, BAT, and FTM, show a slight price increase at UTC hour 20:00, i.e., the closing of NYSE. These two effects are not found in GALA, CHZ, and AXS.

At the opening time of Hang Seng, some of the DeFi Platforms and Tokens and Blockchain-Based Gaming, such as SAND, CRV, GALA, YFI, and AXS, display their all-day high price. Other DeFi and Blockchain-Based Game and Music tokens, such as MATIC, AUDIO, SNX, GALA, COMP, and BAT, display their maximum daily price at UTC hour 8:00, which corresponds to the opening of LSE. Most of the tokens show negative hourly average returns after LSE closes at UTC hour 16:00, except for SNX, YFI, and GALA. GRT and MKR have positive hourly average returns between UTC hours 4:00 to 9:00, i.e., during the second part of the Hang Seng trading day until the opening of LSE. Several tokens, including LINK, UNI, SAND, GALA, MKR, and COMP, attain their top hourly average volumes and volatility during the period of overlap of LSE and NYSE trading between UTC hours 13:00 and 16:00. At UTC hour 14:00, we observe a daily peak in volatility and volume of several tokens. For COMP and MKR, the peak in volume and volatility occurs earlier at UTC hour 13:00. CHZ, AUDIO, and GRT display multiple peaks in volume, not all of them being associated with increased return variances. Similarly, the volatility of SAND and AVE peaks at UTC hour 14:00 (LSE closing), although the all-day maximum volume is observed later in the day at UTC hours 16:00 and 19:00 (NYSE closing), respectively. We observe that the patterns of

[6] LINK, UNI, ENJ, AAVE, CRV, SNX, UMA, AXS, CHZ, MKR, GRT, GALA, BAT

KNC are quite different from those observed in other tokens. For comparison, the intraday patterns of frequently traded tokens are displayed in Figures 37 to 53, (Jasiak and Zhong, 2024, Appendix B).

Figure 2.3: LINK



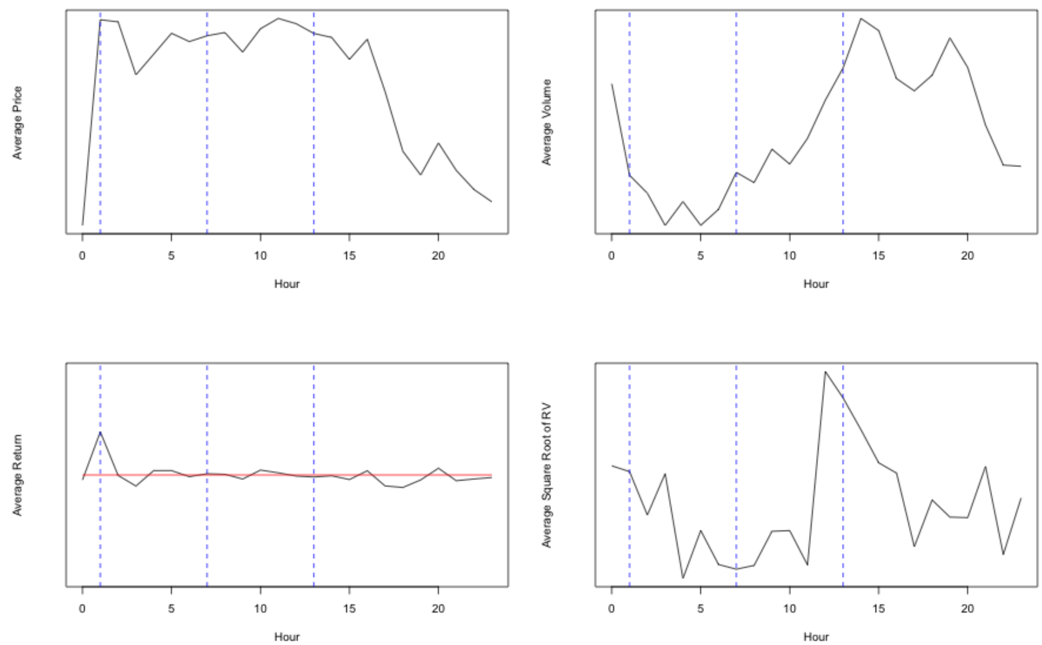
The graph displays for each hour of the day: the average prices of Chainlink (top left), average volumes of Chainlink (top right), average returns of Chainlink (bottom left) and the volatility of Chainlink (bottom right). The vertical lines in the average volumes and log returns plots indicate the 25th and 75th percentiles of sample density. The zero value of average log returns is marked with a red line.

Figure 2.3 shows the average patterns in LINK for illustration. We find statistical evidence of differences between the returns on LINK at different hours. Let us again discuss the statistical evidence of distinct return patterns at UTC hour 1 (opening of Hang Seng), UTC hour 5 (opening of the second part of Hang Seng trading day) and UTC hour 20 (closing of NYSE).

The Mehta and Gurland (1969) test applied to LINK returns provides the evidence against the null hypothesis of equal means of returns at hour 1 compared with hours 0, 3, 4, 5, 6, 7, 9, 10, 11, 13, 16, 20, 21 and 22, at hour 5 compared with hours 1 and 17, and at hour 20 compared with hours 1, 2, 8, 12, 14, 15, 17, 18, 19. The Levene's test for homogeneity of variance reveals statistically significant differences in the variance of LINK returns between hour 1 and hours 4, 5, 6, 7, 8, 9, 10, 11, 12, 17, and 22. Statistically significant differences in the variance are also documented at hour 5 compared with hours 0, 1, 13, 14, 15, and 23, and at hour 20 compared with hours 0, 6, 7, 13, 14, 15, 23. The Kolmogorov-Smirnov test of distribution provides evidence against the null hypothesis of common distribution of LINK returns at hour 1 and hours 3, 4, 5, 6, 7, 8, 9, 10, 11, 16, 20, 21, 22, then hour 5 compared with hours 1, 2, 9, 12, 13, 14, 15, 17, 18, 19, and 23, as well as hour 20 compared with hours 1, 2, 9, 12, 14, 17, 18 and 19.

From the results reported in Sections 3.1.1 and 3.1.2, we conclude that the native cryptocurrencies and frequently traded tokens share several common intraday patterns revealed by the hourly averages. The intraday averages of prices, returns, volatility, and volumes are influenced by the operating times of major stock markets, as illustrated in Figure 2.4.

Figure 2.4: Typical Intraday Patterns of Native Coins and Tokens



The graph displays the typical intraday patterns of native coins and tokens: the average prices (top left), average volumes (top right) average returns (bottom left) and the volatility (bottom right). The red line in average log returns marks a zero value. The openings of Hang Seng, LSE and NYSE are marked by vertical lines.

These typical intraday patterns of price, return, volume, and RV are calculated by averaging the series of prices, volumes, and volatilities of natives and tokens. The returns are calculated from the averaged price series of natives and tokens. The openings of Hang Seng, LSE, and NYSE are marked by the vertical lines in Figure 2.4.

We observe that after UTC hour 14:00, corresponding to the opening of NYSE, most native coins and tokens have lower prices and higher volumes. The volatility spikes shortly before the NYSE opens. The price increase at the opening of Hang Seng at UTC hour 1:00 tends to be associated with positive returns. There is another small positive return at UTC hour 20:00, corresponding to the closure of NYSE.

2.3.1.3 Stablecoins

Let us now describe the intraday patterns in stablecoins and comment on the weak evidence of common patterns.

The stablecoins have lower price volatility than the native cryptocurrencies, as mentioned at the beginning of this section. For example, we can compare the coefficient of variation (CV) of BTC, a native cryptocurrency, with the CV of Tether (UTSD), a stablecoin, by computing the two following statistics:

$$CV_1 = \frac{\sigma}{\bar{x}} \text{ and } CV_2 = \frac{x_{max} - x_{min}}{\bar{x}}$$

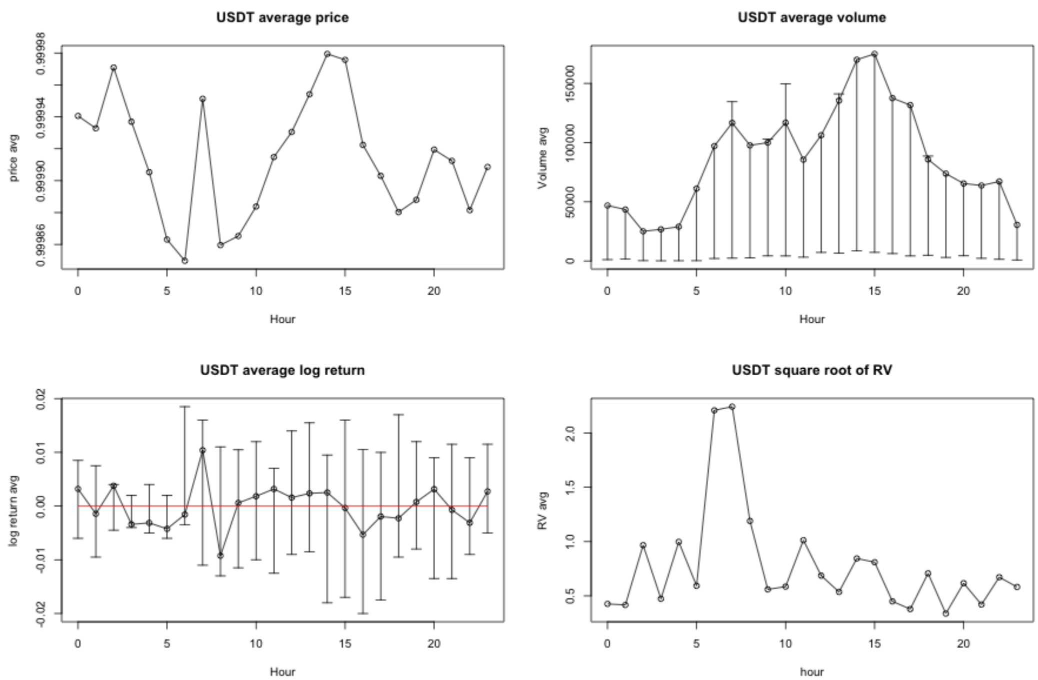
We find that CV_1 is 19.41% for BTC as compared to 0.11% for USDT. Similarly, CV_2 is 81.41% for BTC and 3.91% for USDT.

The prices of stablecoins do not share much common periodic patterns, unlike the native cryptocurrencies and tokens. Moreover, the intraday changes in prices and returns can be very small and need to be interpreted with caution.

The hourly average price of USDT is high after the opening of Hang Seng at UTC hour 2:00 and one hour before Hang Seng closes at UTC hour 7:00. Later on, the USDT price rises again after the opening of NYSE to its daily maximum between hours 14:00 and 15:00. The average return of USDT is positive and at its all-day high at UTC hour 7:00, before the closing of Hang Seng. It remains positive between UTC hours 9:00 to 14:00 when LSE is operating. The hourly average volume of USDT is high during the LSE trading day between UTC hours 8:00 and 14:00 and peaks at the opening hour of NYSE at UTC hour 14:00. The hourly average RV of USDT is high between UTC hours 6:00 and 7:00. For comparison, the hourly average price of USDC (Figure 54, (Jasiak and Zhong, 2024, Appendix B)) attains its top level at UTC hour 0:00, i.e. one hour before the opening of Hang Seng. The hourly average volume of USDC is high at UTC hour 10:00 and after the closure of NYSE at UTC hour 20:00. The hourly average RV of USDC peaks between UTC hours 13:00 and 14:00

when LSE is operating, and NYSE is about to open. That RV peak is associated with low prices and positive returns at UTC hour 14:00 when NYSE opens.

Figure 2.5: USDT



The graph displays for each hour of the day: the average prices of Tether (top left), average volumes of Tether (top right), average returns of Tether (bottom left) and the volatility of Tether (bottom right). The vertical lines in the average volumes and log returns plots indicate the 25th and 75th percentiles of sample density. The zero value of average log returns is marked with a red line.

Figure 2.5 illustrates the behavior of USDT. We find less statistically significant differences across the UTC day than in the natives and tokens.

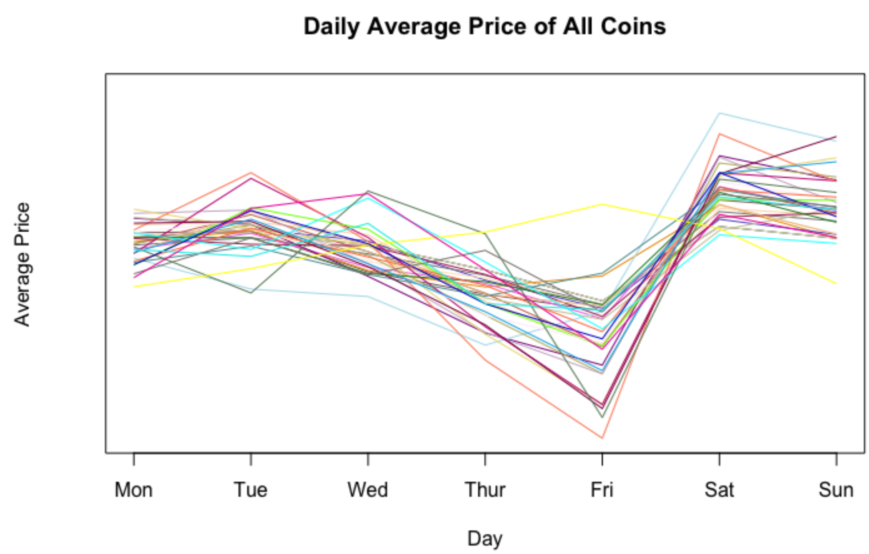
The Mehta and Gurland (1969) tests of equal means of USDT returns provide evidence against the null hypothesis of equal means only when the USDT returns at hour 5 are compared with hour 0, and at hour 20 compared with hours 8 and 16. The Levene's test for homogeneity of variance provides evidence against the null hypothesis of equal variances of USDT returns between hours 1 and 15, hour 5 compared with 14, as well as hour 5 compared with 15. There is no evidence against the null hypothesis of equal variance of returns on

USDT at hour 20 and any other hours. The Kolmogorov-Smirnov test of distribution provides evidence against the null hypothesis of common distribution of returns on USDT at hour 1 compared with hour 6, hour 5 compared with hours 6, 7, 12, 13, 14, 15, 16, 18, 19 and 23, as well as hour 20 compared with hours 3, 4, 6 and 23. We find that there are less statistically significant intraday patterns in the returns of USDT than in the returns on Bitcoin and LINK.

2.3.2 Intraweek Patterns

This Section examines the intraweek patterns in native cryptocurrency, tokens and stablecoins. Our analysis is motivated by the price behavior illustrated in Figure 2.6, where the curves represent the prices of all frequently traded natives and tokens rescaled by their own weekly price averages plotted over 7 days of the week.

Figure 2.6: Daily Average (Rescaled) Prices of Natives Coins and Tokens (No Stablecoins)



All coin's daily average prices are rescaled (but not demeaned) by their own averages, e.g., BTC's daily average prices are rescaled (but not demeaned) by the prices of BTC on days 1 to 7 (Monday to Sunday)
We observe a drop in prices on Fridays and higher prices over the weekends.

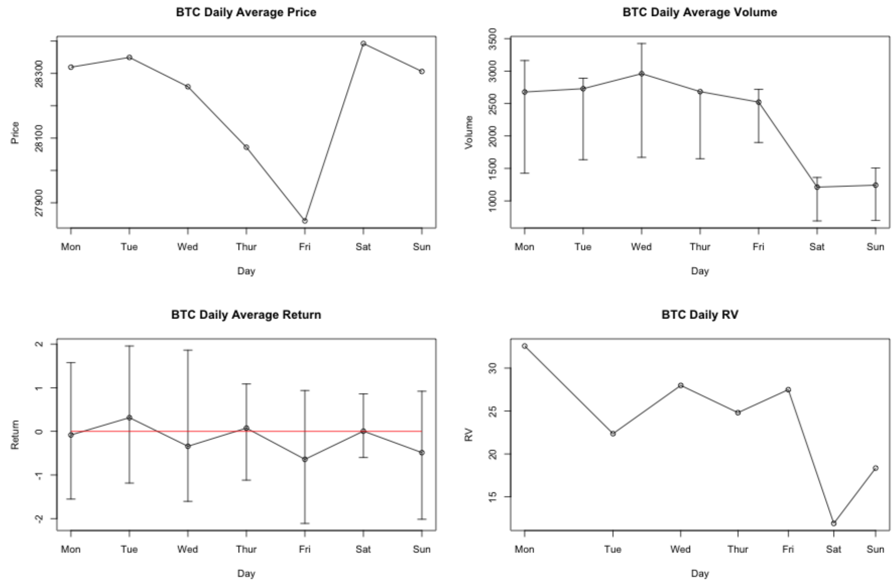
We observe that the coin prices are, on average, relatively steady during the working days and decrease on Fridays. The highest average prices tend to be recorded on Saturdays. The stablecoin prices appear constant because of their low volatility compared to the natives and tokens. For this reason, they are not reported in Figure 2.6.

2.3.2.1 Native Cryptocurrency

Let us now describe the common intraweek patterns in the prices, returns, volumes and volatility of native cryptocurrency.

We observe a common pattern in the intraweek average prices of native cryptocurrency, with daily averages decreasing gradually over the working days down to their lowest level on Fridays. The daily average prices of BTC and other native coins rebound on the weekends and tend to be high on Saturdays. The daily average returns tend to be positive on Tuesdays when most natives also show low average volatility. The highest volumes are mostly recorded on Wednesdays or Thursdays. The average volatility and volumes are low on the weekends. There are some exceptions, however. For example, XRP displays its lowest price on Thursday instead of Friday. The volume of ETH and ADA peak on Tuesdays, and the RV of SGB is at its highest level on Fridays. The intraweek patterns of native cryptocurrencies are displayed in Figures 28 to 36, (Jasiak and Zhong, 2024, Appendix B).

Figure 2.7: BTC



The graph displays for each day of the week: the average prices of Bitcoin (top left), average volumes of Bitcoin (top right), average returns of Bitcoin (bottom left) and the volatility of Bitcoin (bottom right). The vertical lines in the average volumes and log returns plots indicate the 25th and 75th percentiles of sample density. The zero value of average log returns is marked with a red line.

As an illustration, Figure 2.7 displays the intraweek patterns of Bitcoin. We confirm the intraweek results of Hansen et al. (2024) who show that the volume of Bitcoin and Ether are close over the week and lower on the weekends. However, we observe that RV does not follow the same pattern and is higher on Monday. Regarding Ma and Tanizaki (2019), their evidence is mixed. The linear model reveals that Bitcoin has a significant non-zero return only on Monday. Their two SV models reveal that Bitcoin has significant non-zero returns on Tuesday and Friday, and the return on Monday is no longer significant. In our sample, BTC has the highest returns on Tuesday and the lowest returns on Friday. In terms of volatility, Ma and Tanizaki (2019) found that Monday and Thursday volatilities are significantly different from other days, which is partly consistent with our results.

The Levene's test for variance homogeneity provides us the evidence against the null hypothesis of equal variances of daily Bitcoin returns compared between Monday and Saturday, Sunday, Tuesday and Saturday, as well as Wednesday and Saturday, Wednesday and Sunday, Thursday and Saturday, Friday and Saturday, Saturday and Sunday. The Kolmogorov-Smirnov test provides the evidence against the null hypothesis that "the two samples were drawn from the same distribution" when Bitcoin returns are compared between Tuesday and Saturday, Wednesday and Saturday, and Friday and Saturday.

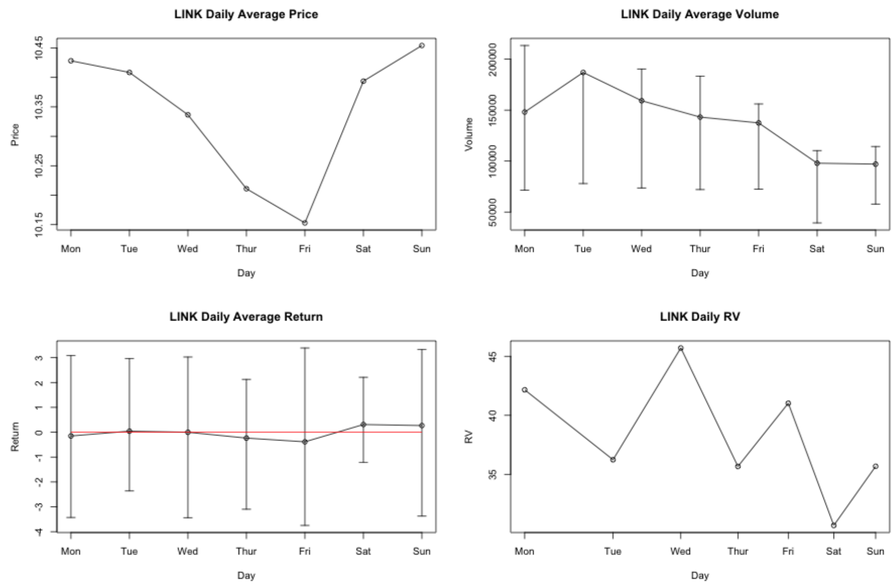
2.3.2.2 Tokens

There exist common patterns in the intraweek dynamics of tokens, which are summarized below.

Several tokens display similar common intraweek patterns revealed by the daily average prices, which decrease over the week to their lowest level on Friday and jump up on the weekends. CHZ is the only token in our sample that has the lowest average price on Thursdays. Like in the intraday data, KNC also shows a different price pattern. Its price increases during the week, Monday through Friday, and falls on the weekends, Saturday through Sunday. The daily average volumes and volatilities of tokens decrease throughout the week as well and are at their lowest levels over the weekends. The daily average returns display low volatility and tend to be positive on Saturdays. The intraweek patterns of tokens are displayed in Figures 64 to 85, (Jasiak and Zhong, 2024, Appendix B).

As an example, Figure 2.8 illustrates the intra-week pattern of Chainlink (LINK).

Figure 2.8: LINK

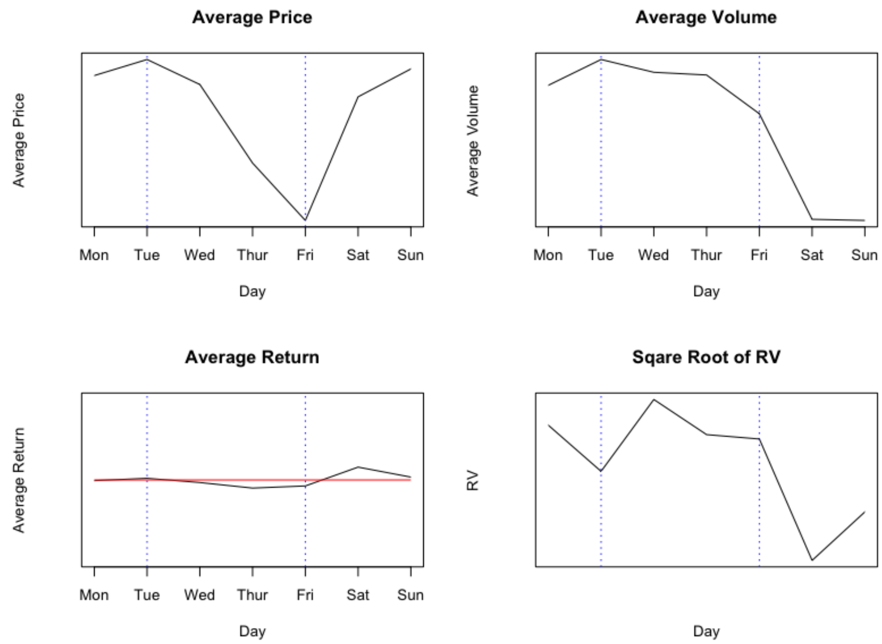


The graph displays for each day of the week: the average prices of Chainlink (top left), average volumes of Chainlink (top right), average returns of Chainlink (bottom left) and the volatility of Chainlink (bottom right). The vertical lines in the average volumes and log returns plots indicate the 25th and 75th percentiles of sample density. The zero value of average log returns is marked with a red line.

The Kolmogorov-Smirnov test provides the evidence against the null hypothesis "the two samples were drawn from the same distribution" for the returns on LINK compared between Tuesday and Saturday, Wednesday and Saturday and Friday and Saturday.

We summarize the results of Sections 3.2.1 and 3.2.2 by displaying the intraweek patterns common to both native cryptocurrencies and tokens. Figure 2.9 shows a typical intraweek pattern of prices, returns, volumes and volatility of native coins and tokens.

Figure 2.9: Typical Intraweek Patterns of Native Coins and Tokens



The graph displays the typical intraweek patterns of native coins and tokens: the average intraweek prices (top left), average intraweek volumes (top right) average intraweek returns (bottom left) and the intraweek volatility (bottom right). The red line in average log returns marks a zero value. On Tuesdays and Fridays (marked with vertical lines), we observe maximum and minimum intraweek average prices, and kinks in the average returns, volume and volatility curves.

The typical intraweek patterns of price, return, volume, and RV are calculated by averaging the series of daily prices, returns volumes, and RV over the native cryptocurrencies and tokens [7]. The vertical lines indicate Tuesdays and Fridays associated with the maximum and minimum intraweek average prices, respectively. We observe that the daily average prices decrease over the week to attain their lowest levels on Fridays. The weekends are characterized by low average volumes and low volatility.

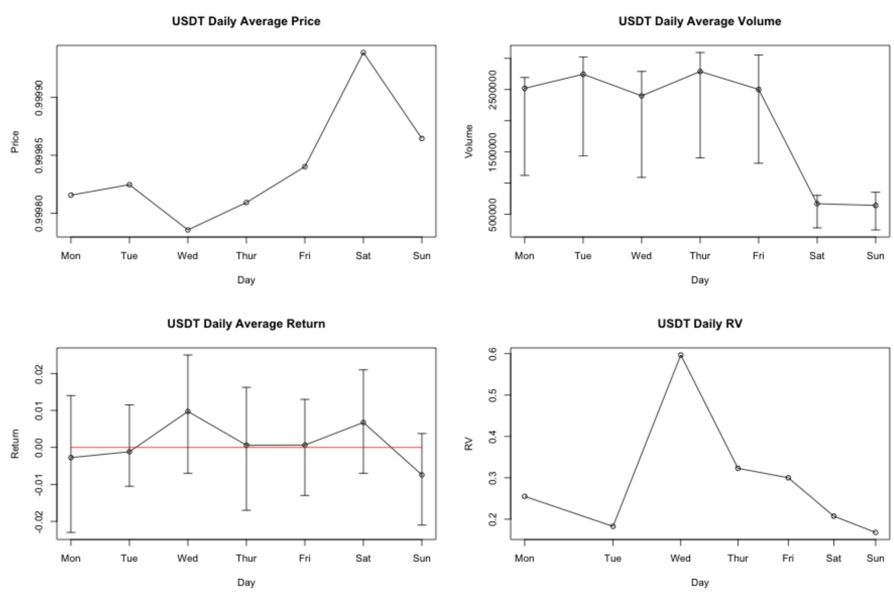
[7]The daily average returns plotted for each individual coin in Figure 2.7, Figure 2.8, and in Figures 55 to 87, (Jasiak and Zhong, 2024, Appendix B), are calculated by removing returns exceeding the mean by two standard deviations in absolute value.

2.3.2.3 Stablecoins

Let us now examine the intraweek dynamics of stablecoins to find out if they display common patterns.

Stablecoins share few common intraweek patterns. The daily average prices of stablecoins tend to be lower in the middle of the week. They attain their weekly maximum at the end of the week. Figure 2.10 shows the intraweek patterns of Tether (USDT).

Figure 2.10: USDT



The graph displays on each day of the week: the average prices of Tether (top left), average volumes of Tether (top right), average returns of Tether (bottom left) and the volatility of Tether (bottom right). The vertical lines in the average volumes and log returns plots indicate the 25th and 75th percentiles of sample density. The zero value of average log returns is marked with a red line.

The daily average volumes of USDT and USDC tend to decrease over the working days down to their lowest levels on the weekends. The intraweek patterns of USDC are given in Figure 86, (Jasiak and Zhong, 2024, Appendix B). The intraweek patterns of DAI are quite distinct from USDT and USDC. For comparison, they are displayed in Figure 87, (Jasiak and Zhong, 2024, Appendix B).

We find more evidence of statistically significant differences in USDT across the week than across the day. The Mehta and Gurland (1969) test for equal means of USDT daily returns provides evidence against the null hypothesis when the daily returns are compared between Monday and Saturday, Tuesday and Sunday and Saturday and Sunday.

The Levene's test for homogeneity of variance reveals evidence against the null hypothesis of equal variances when we compare the variance of daily returns between Monday and Tuesday, Monday and Sunday, Tuesday and Thursday, Tuesday and Friday, Wednesday and Sunday, Thursday and Saturday, Thursday and Sunday, and Friday and Sunday.

The Kolmogorov-Smirnov test for equal distribution provides evidence against the null hypothesis when we compar the returns between Tuesday and Saturday, Wednesday and Saturday and Friday and Saturday.

2.4 Conclusion

Common periodic intraday and intraweek patterns are detected in prices, returns, volumes and volatility of native cryptocurrency and tokens. Generally, investors can benefit from lower prices by trading between 4 pm and 10 pm Eastern Standard Time, i.e. between the closing time of NYSE and the opening time of Hang Seng, and on Fridays. Stablecoins have distinct intraday and intraweek dynamics.

Chapter 3

Functional CAPM Analysis in Cryptocurrency Markets

Coauthorship Statement

This chapter is based on joint work titled "Intraday and Daily Dynamics of Cryptocurrency" with Dr. Joann Jasiak. All authors contributed equally to the conceptualization, data preparation, empirical modeling, analysis, and writing of the manuscript.

A version of this chapter is currently accepted and published at `International Review of Economics and Finance`.

3.1 Introduction

In the existing literature, the validity of a traditional one-factor daily CAPM applied to daily cryptocurrency returns was documented by Tavares et al. (2021). Their results show sizable betas with small and statistically insignificant alphas. Tavares et al. (2021) conducted the CAMP analysis of the ten highest market capitalization cryptocurrencies as the assets (without stablecoins) with the capitalization-weighted price index CRIX as the market factor. They used data from www.CoinMarketCap.com from July 2014 to October 2017. During that period, Bitcoin was the dominant digital asset, and the majority of coins included in the CRIX index were natives. Stablecoins were not included in the index during that period. More recently, Dunbar and Owusu-Amoako (2022) estimated a CAPM for Bitcoin, Ether and Ripple with their own value-weighted cryptocurrency market portfolio, excluding stablecoins, from daily returns over the period 2014-2021. The inference from weekly data was considered in Shen et al. (2020), who proposed a three-factor pricing model, considering size, market and reversal factors. They claimed that their model outperforms the excess market return C-CAPM. Shahzad et al. (2021) proposed a four-factor pricing model by adding contagion risk to the three-factor model of Shen et al. (2020) and showed that bad contagion should be considered a risk factor while modelling the price of the cryptocurrency. The periodic patterns were not considered in these models.

We also document several additional results based on the analysis of the correlation matrix. In particular, stablecoin returns are found to be uncorrelated with one another or with the returns on other types of cryptocurrencies. The connectedness measures (Billio et al., 2012) indicate that native cryptocurrencies and tokens are systemically connected, and the tokens contribute more to the systemic risk than native coins. Stablecoins seem to be immune to the systemic risk associated with the cryptocurrency market.

Our second contribution is in introducing the functional Capital Asset Pricing Model for cryptocurrency as an extension of the CAPM, to accommodate periodic patterns by replacing the time series of returns with the series of return functions over either 24 hours or 7 days, for intraday and daily data, respectively. The estimation of the functional CAPM shows that an affine relationship between the functions of returns on a cryptocurrency and the market portfolio exists for several cryptocurrencies, including Bitcoin and Ether among the natives, and LINK among the tokens. We document periodic variations in the functional betas during the day, which are influenced by the openings and closures of major stock markets. In daily data, we observe slight variations of functional betas across the week.

To estimate the functional CAPM, we approximate the cryptocurrency market portfolio by applying the principal component analysis (PCA) to the intraday and intraweek cross-sectional return correlation matrices of cryptocurrencies following the approach of Avellaneda et al. (2022). The PCA reveals one common factor in the cryptocurrency market in hourly and daily data, to which stablecoins do not contribute. The returns on the cryptocurrency market portfolio display periodic patterns comparable to those in the SPCBDM index, launched in July 2021 and recorded daily, Monday through Friday.

3.2 Methodology

The seasonal patterns characterize not only the individual cryptocurrency but also the portfolios of cryptocurrency. To study these patterns, we use the Principal Component Analysis (PCA) to build a cryptocurrency market portfolio and conduct inference on the connectedness of coins. We also describe the functional regression, which allows us to estimate a relationship between the returns on the coins and the market portfolio.

3.2.1 Principal Component Analysis (PCA)

The PCA is a decomposition of the variance-covariance matrix (variance matrix, henceforth) of a multivariate random variable. For matrix X of dimension $T \times \mathcal{N}$ with columns containing the random variables denoted by $x_i, i = 1, ..., \mathcal{N}$ of length T, the theoretical variance matrix is $E[(X - E(X))(X - E(X))']$, where $E(X)$ is the vector of expected values. In practice, the expected values are estimated from the sample averages and the variance matrix is estimated from the sample of T demeaned observations.

Alternatively, the PCA can be applied to the variance matrix of standardized random variables (correlation matrix) $\tilde{X}'\tilde{X}$, containing the columns $\tilde{x}_i/\sigma(\tilde{x}_i)$, where $\sigma(\tilde{x}_i)^2$ is the variance of component $\tilde{x}_i$.

The PCA is a system of linear combinations of the variables (James et al., 2021). The first principal component is the linear combination of all variables that explains the highest

percentage of the total variance. The remaining linear combinations are orthogonal to the first principal component and explain lower percentages of the variance.
The PCA applied to the symmetric matrix $\tilde{X}'\tilde{X}$ is:

$$\tilde{X}'\tilde{X} = Q\Lambda Q' = \sum_{i=1}^{\mathcal{N}} \lambda_i q_i q_i'.$$

where Q is an orthogonal $\mathcal{N} \times \mathcal{N}$ matrix whose columns $q_i, i = 1, ..., \mathcal{N}$ are the eigenvectors of $\tilde{X}'\tilde{X}$ and Λ is a diagonal matrix whose elements are the eigenvalues of $\tilde{X}'\tilde{X}$.

In our application to cryptocurrency data, the PCA decomposes the correlation matrix of daily and hourly returns into orthogonal common factors (principal components) with respect to their decreasing explanatory power.

3.2.2 Connectedness Measure

The inter-connectedness of cryptocurrencies can be assessed by using the connectedness measure of Billio et al. (2012) based on the PCA applied to the variance-covariance matrix of $\mathcal{N}$ demeaned and standardized asset returns (correlation matrix) [1]. This measure provides us with the exposure and contribution of each cryptocurrency to the cryptocurrency market risk, which can be interpreted as the systemic risk. Billio et al. (2012) compare the total risk of the system defined by $\sum_{k=1}^{\mathcal{N}} \lambda_k$ with the risk associated with the first n principal components, defined by $\sum_{k=1}^{n} \lambda_k$. This ratio is called the cumulative variance percentage and is defined as:

$$h_n = \frac{\sum_{k=1}^{n} \lambda_k}{\sum_{k=1}^{\mathcal{N}} \lambda_k}$$

where h_n denotes the cumulative variance percentage explained by the first n principal components. The system is highly interconnected if a small number of n principal components can explain most of the total risk. The higher the fraction of total variance explained by the

[1]We apply the PCA to demeaned and rescaled returns in Section 5.

first few common factors for a group of cryptos, the more interconnected the cryptos in that group. Therefore, a larger return connectedness implies a greater systemic risk among the cryptos.

Billio et al. (2012) measure the connectedness in terms of individual exposure and contribution to the system, denoted by $PCAS_{i,n}$. It is the contribution of asset i to the risk of the system.

$$PCAS_{i,n} = \sum_{k=1}^{n} \frac{1}{2} \frac{\sigma_i^2}{\sigma_T^2} Q_{ik}^2 \lambda_k$$

where, σ_i^2 is the variance of asset i, and σ_T^2 is the total variance of the system $\sigma_T^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sigma_i \sigma_j Q_{ik} Q_{jk} \lambda_l$ where, N is the number of assets. Q_{ik} represent the factor loading of i in the k^{th} eigenvector and and λ_k is the k^{th}eigenvalue of the correlation matrix. n is determined by the threshold of cumulative variance explained.

3.2.3 Cryptocurrency Eigenportfolio

To approximate the cryptocurrency market portfolio, we build a principal eigenportfolio from the first eigenvector of the variance-covariance matrix of demeaned and standardized returns, following the approach of Avellaneda et al. (2022). It is widely recognized that the principal eigenportfolio is a good proxy for the capitalization-weighted market portfolio (Avellaneda et al., 2022). We extend this concept to the cryptocurrency market with $\mathcal{N}$ assets, and denote the first principal eigenportfolio by E^1, where

$$E^1 = \frac{1}{c} s^{-1} q_1$$

and

$$s = diag(s_1, ..., s_N)$$

where $s_i = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_i(t) - \bar{r}_i)^2}$, $r_i(t)$ is the return on asset i at time t, q_1 is the first eigenvector and $c = \sum_{i=1}^{\mathcal{N}} q_{1,i}/s_i$. Note that the eigenportfolio is normalized to sum to one by using c. The return of the principal eigenportfolio at time t is denoted by

$$f(t) = \sum_{i=1}^{\mathcal{N}} r_i(t) E_i^1.$$

and approximates the return on the capitalization-weighted market portfolio (Avellaneda et al., 2022).

3.2.4 Functional Regression

The functional regressions reveal the relationships between the functions of 24 hourly (excess) returns on a given cryptocurrency, as the dependent variable, and the functions of 72 hourly (excess) returns on the cryptocurrency market portfolio (eigenportfolio) as the regressor (Ramsay and Silverman, 2005a). More specifically, we consider the excess returns and estimate a functional model:

$$y_d(s) = \int_{-24}^{48} \Pi(s,t) x_d(t) dt + u_d(s),$$

where $y_d(s) = r_d(s) - r_{0,d}(s)$ is the excess return of time $s \in S$ of day d on a cryptocurrency, $x_d(t)$ is the excess market return on day d at time t, where $x_d(t) = f_d(t) - r_{0,d}(t)$, $u_d(s)$ is a zero-mean error term, $\Pi(s,t)$ is the integral operator, and r_0 is the risk-free rate. The integral operator Π depicts the dynamic relation between the returns on a specific cryptocurrency and the cryptocurrency market factor. By integrating over the range -24 to 48, we examine the functional relationship over the previous, current and next days.

The daily functional regression of the return function on the market factor portfolio function is:

$$y_i(s) = \int_{-7}^{7} \Pi(s,t) x_i(t) dt + u_i(s)$$

where $y_i(s)$ is the excess return of day $s \in S$ of week i on a cryptocurrency, $x_i(t)$ is the excess market return of week i (the week starts on Monday, i.e., 1 to 7 represents Monday to Sunday respectively) at day t, and $u_i(s)$ is a zero-mean error term. By integrating over the range of -7 to 7, we examine the functional relationship over the previous and current week.[2]

Benatia et al. (2017) shows that the estimator of $\Pi(s,t)$ calculated from discrete data can be written as

$$\hat{\Pi}_\alpha(s,t) = \frac{1}{N}\underline{y}(s)'(\alpha I + M)^{-1}\underline{x}(t)$$

where N is the number of days in the sample, $\underline{y}$ and $\underline{x}$ are the $N \times 1$ vectors with ith element $y_i(s)$ and $x_i(t)$ respectively. The matrix M is of dimension $N \times N$ with (l,i) element $\langle x_l, x_i \rangle / N$. The regularization parameter α is chosen by leave-one-out cross-validation (LOOCV).

The variance of the integral operator $\hat{\Pi}(.)$ for discrete data can be estimated as follows:

$$V\hat{\pi}_u(s) \simeq \left(\sum_{i=1}^{N} x_i x_i' + \alpha_u I\right)^{-1} \sum_{i=1}^{N} x_i x_i' \hat{u}_i^2(s) \left(\sum_{i=1}^{N} x_i x_i' + \alpha_u I\right)^{-1}$$

This expression is used to assess the statistical significance of coefficients in the daily functional regressions. In intraday regressions, we use the cut-off points of estimated coefficients to determine the statistical significance because of sparsity and singularity issues in high dimensional data.

3.3 Cryptocurrency Market Portfolio

In this section, we build the cryptocurrency market portfolio and compare it with the cryptocurrency index SPCBDM to determine if it approximates the capitalization-weighted index well. We want to find out if all types of cryptocurrencies contribute to the market portfolio

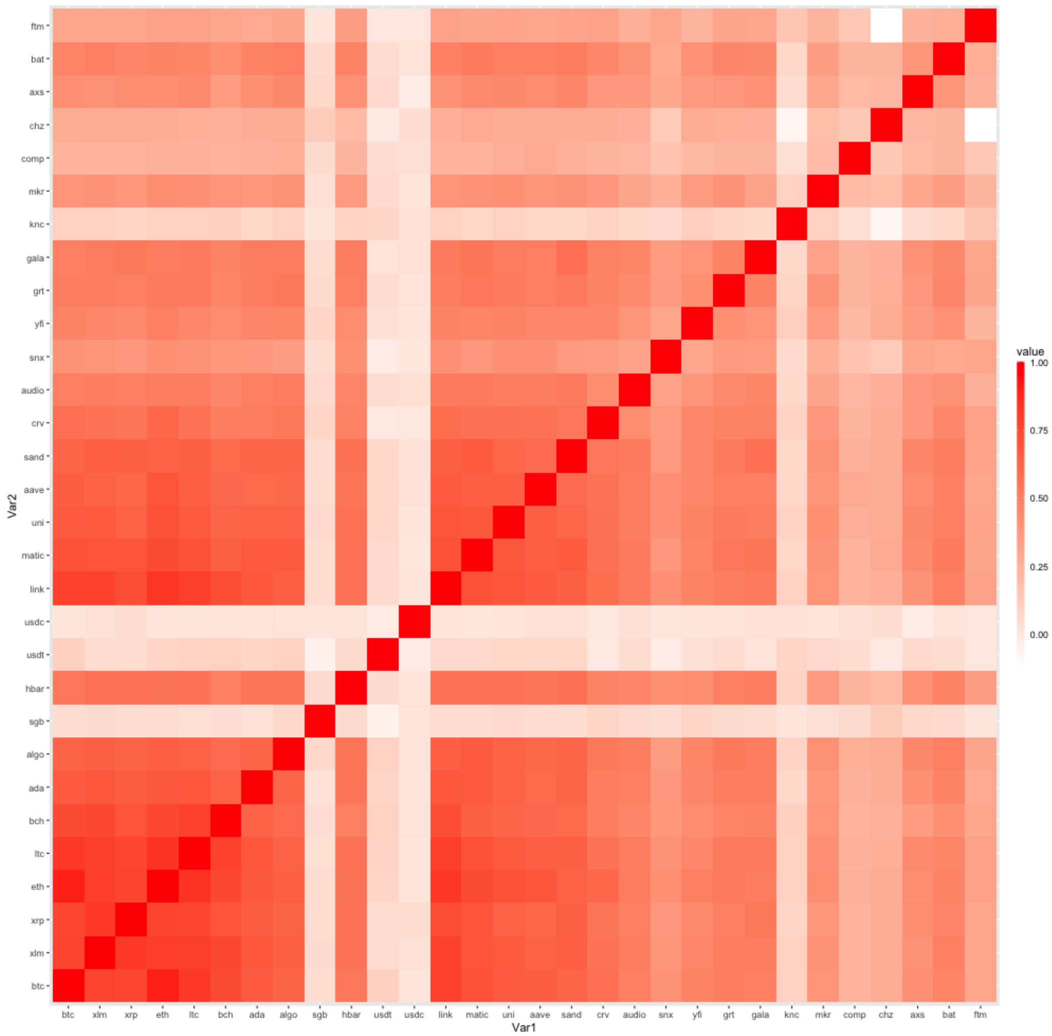
[2]The future 7 days are omitted as no statistically significant effects were found in empirical research

and if the daily returns on the market portfolio display similar dynamics and periodicity as the SPCBDM. The allocations of the market portfolio are found by applying the PCA to the matrix of return correlation displayed in the next Section.

3.3.1 Correlation of cryptocurrency returns

The correlation matrix is a basic measure of the cross-sectional dependence of returns on coins over the sampling period. The contemporaneous cross-correlation analysis of hourly returns is applied to 30 frequently traded coins. Figure 3.1 shows the heatmap of hourly return correlations of these cryptocurrencies.

Figure 3.1: The Heatmap of Hourly Return Correlation



A darker colour indicates that the two coins are more correlated. A lighter colour indicates that the two coins are less correlated. We observe that the stablecoins are uncorrelated with other coins and with one another.

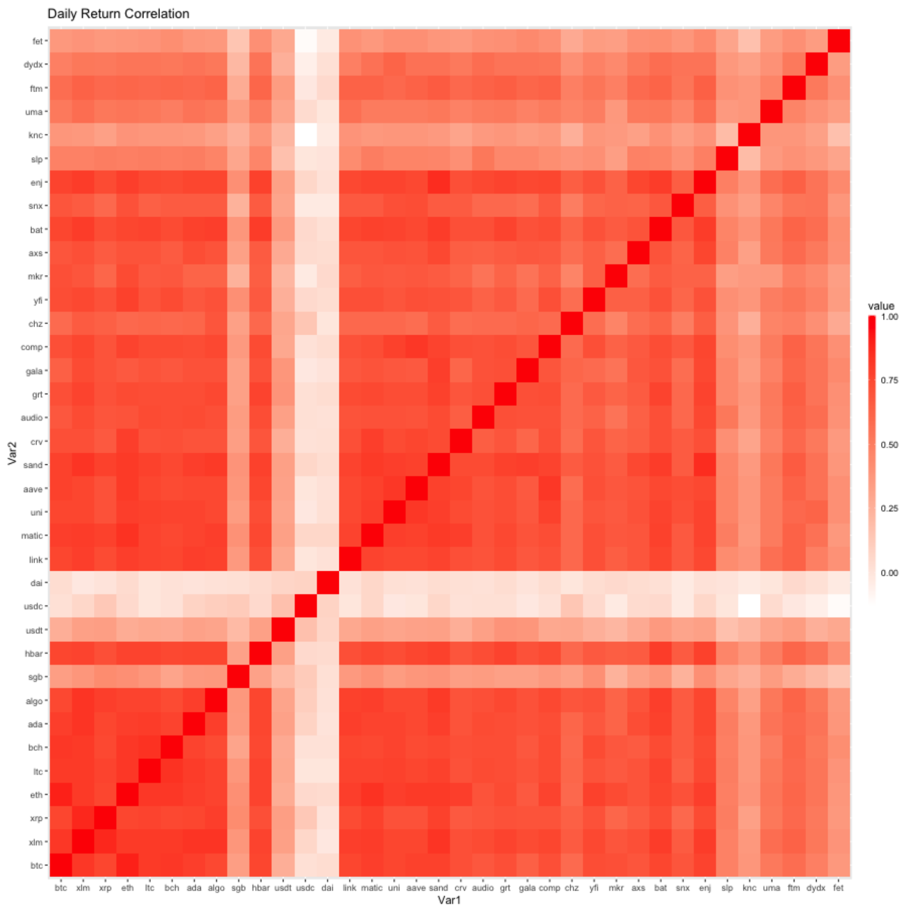
The axes of Figure 3.1 show the symbols of natives, stables and tokens, displayed in this order. Starting from the origin, the coins in each group are ranked with respect to the percentages of hourly non-zero volumes, which are decreasing from left to right. Hence, the coins located on the right side of the figure are traded less often than other coins and have more zero values of returns, which can bias the correlations down and may explain why they appear less correlated with the remaining coins. The dark color represents a strong correlation, while the light color depicts correlations close to 0. The return correlations are positive and high, especially for the natives and some tokens. The exceptions are SGB (native coin) and KNC (token), which are not highly correlated with other coins [3]. The returns on stablecoins USDT and USDC show little correlation with the returns on other coins (columns 11 to 12 on the horizontal axis). These correlations are close to zero and appear almost white in Figure 3.1.

These results can be compared with cross-sectional correlations of daily cryptocurrency returns. These are computed from 36 frequently traded coins, because more coins satisfy the minimum transactions condition. The correlation heatmap of daily returns is displayed in Figure 3.2.

We find that the daily returns of native coins and tokens are also positively and highly correlated, except for SGB. Moreover, there is no evidence of return correlation in stablecoins in columns 11 to 13 (USDC, USDT and DAI), which is consistent with the intraday results. The correlation between the returns on stablecoins and other cryptocurrencies (i.e., natives and tokens) is close to 0 too.

[3] Recall that KNC displays distinct periodic patterns as pointed out in Section 4.1.2.

Figure 3.2: The Heatmap of Daily Return Correlation



A darker colour indicates that the two coins are more correlated. A lighter colour indicates that the two coins are less correlated. We observe that the stablecoins are uncorrelated with other coins and with one another.

The daily returns on KNC show more correlation with the returns on other coins than in the hourly correlation matrix, as indicated by the darker shade in the heatmap.

3.3.2 PCA analysis of returns

In this Section, we document one factor ion the cryptocurrency market and explore the connectedness measures of coins.

The PCA is applied to the intraday correlation matrix of dimension 4296 by 30, computed from the demeaned and rescaled returns on one particular coin [4]. Table A.3 in Appendix A displays the eigenvalues associated with each principal component. We observe that the eigenvalues cut-off, i.e. the first one is 10 times larger than the following one, suggesting the existence of one market factor. Next, we calculate the variance percentage explained by each principal component, $h_i = \lambda_i / \sum_1^{30} \lambda_i$, and the cumulative variance percentage explained by the first n principal components, $h_n = \sum_1^n \lambda_i / \sum_1^{30} \lambda_i$. The first principal component captures 44.55% of the variability of hourly returns. The first 10 (resp. 20) principal components explain 73.83% (resp. 91.40%) of hourly return variation.

Similar results pointing to the existence of one market factor are obtained from the daily data. The PCA is applied to the variance matrix of daily returns of dimension 364 by 36, where each column is demeaned and rescaled by dividing the demeaned columns by the standard deviations. The results of PCA analysis of daily returns are given in Table A.4 of Appendix A. The first eigenvalue is 16 times larger than the following one. Hence, we observe a cut-off point in the eigenvalues and conclude that there is one cryptocurrency market factor. This result is consistent with (Dunbar and Owusu-Amoako, 2022; Tavares et al., 2021). For comparison, the first 10 (resp. 20) principal components explain 81.36% (resp. 92.18%) of the daily return variation.

Table 3.1 displays the hourly and daily connectedness measures $PCAS_{i,n}$ for n = 20.

[4] The row dimension 4296 is 179 days times 24 observations per day. Each column of the matrix is demeaned by subtracting the global mean and rescaled by dividing the demeaned columns by their standard deviations.

As mentioned earlier, the $PCAS_{i,n}$ of an individual coin i illustrates its risk contribution or exposure to the variance (risk) of the system (Billio et al., 2012). We notice that the stablecoins (USDT and USDC) have a much lower contribution to the system compared to natives and tokens. This is consistent with the objective of stablecoins, as they are supposed to remain strongly correlated with the US Dollar to maintain the peg rather than with the cryptocurrency market. Hence, the stablecoins seem immune to the cryptocurrency market risk. We find that the results from daily data are similar to those obtained from the hourly analysis. The stablecoins have much less risk contribution compared to the natives and tokens. SGB has the largest risk contribution/exposure to the daily cryptocurrency system among the natives. We observe that, like in the hourly data, KNC also has the largest risk contribution among the tokens in daily data.

We also find that Bitcoin and the natives, in general, have lower contributions to the risk of the system than the frequently traded tokens. SGB has the highest PCAS (0.0026) among the natives, and KNC has the highest PCAS (0.0050) among the tokens.

Table 3.1: $PCAS$ for Each Cryptocurrency

Coin	Hourly $PCAS_{i,n} * 10^{-2}$	Daily $PCAS_{i,n} * 10^{-3}$
Natives		
BTC	0.091	0.1200
XLM	0.120	0.1200
XRP	0.110	0.1700
ETH	0.120	0.1800
LTC	0.120	0.1700
BCH	0.110	0.1700
ADA	0.130	0.1800
ALGO	0.150	0.1900
SGB	0.260	0.2800
HBAR	0.180	0.1700
Stablecoins		
USDT	0.009	0.0021
USDC	0.007	0.0140
DAI	NA	0.0140
Tokens		
LINK	0.150	0.1900
MATIC	0.140	0.2300
UNI	0.140	0.2000
AAVE	0.200	0.2300
SAND	0.160	0.2200
CRV	0.220	0.2800
AUDIO	0.250	0.2400
SNX	0.270	0.3100
YFI	0.180	0.2200
GRT	0.220	0.2300
GALA	0.240	0.2600
KNC	0.500	0.3400
MKR	0.230	0.2100
COMP	0.360	0.2300
CHZ	0.310	0.2800
AXS	0.240	0.2500
BAT	0.190	0.1800
FTM	0.320	0.3200
UMA	NA	0.2800
DYDX	NA	0.3100
FET	NA	0.3100
ENJ	NA	0.2100
SLP	NA	0.3300

The stablecoins do not contribute to the market risk. Native coins contribute less than the tokens.

As mentioned earlier, the first principal component associated with the largest eigenvalue explains the highest percentage of the system's variance. Hence, the coefficients of the first principal component are used to build the eigenportfolio, which approximates the capitalization-weighted cryptocurrency market portfolio (Avellaneda et al., 2022). The allocations of that eigenportfolio are the coin i's value of the first eigenvector (principal component coefficients) divided by the standard deviation of its return.

The allocations of the hourly and daily eigenportfolio are displayed in Table 3.2 below. The stablecoins whose returns are uncorrelated with the returns on other coins do not contribute to the risk of the cryptocurrency market. Hence, they are removed from the eigenportfolio,

Table 3.2: Allocation of Hourly and Daily Eigenportfolio (without Stablecoins)

Coin	Hourly Allocation	Daily Allocation
Natives		
BTC	0.0807	0.0559
XLM	0.0615	0.0476
XRP	0.0599	0.0411
ETH	0.0636	0.0424
LTC	0.0598	0.0412
BCH	0.0543	0.0414
ADA	0.0453	0.0390
ALGO	0.0424	0.0361
SGB	0.0027	0.0141
HBAR	0.0381	0.0398
Token		
LINK	0.0472	0.0351
MATIC	0.0434	0.0312
UNI	0.0439	0.0335
AAVE	0.0356	0.0297
SAND	0.0370	0.0335
CRV	0.0319	0.0251
AUDIO	0.0257	0.0262
SNX	0.0193	0.0208
YFI	0.0336	0.0294
GRT	0.0306	0.0286
GALA	0.0257	0.0236
KNC	0.0024	0.0122
MKR	0.0228	0.0308
COMP	0.0096	0.0310
CHZ	0.0117	0.0225
AXS	0.0231	0.0277
BAT	0.0337	0.0369
FTM	0.0132	0.0205
UMA	NA	0.0204
DYDX	NA	0.0186
FET	NA	0.0140
ENJ	NA	0.0346
SLP	NA	0.0156

Natives have relatively larger allocations (except SGB) than Tokens.

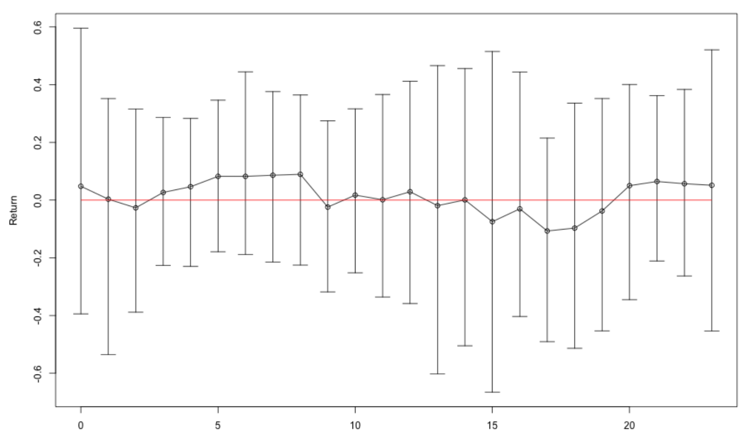
like in Tavares et al. (2021) and Dunbar and Owusu-Amoako (2022). The eigenportfolio computed from daily data contains 5 more coins than the one computed from hourly data: FET, ENJ, DYDX, UMA and SLP. Hence, the allocations of the 30 initial coins are lower on a daily basis, except for native SGB and tokens MKR, KNC and COMP.

3.3.3 Market Portfolio Periodicity

It is interesting to explore if the returns on the market portfolios computed from our hourly and daily data display periodicity, and if the returns on the SPCBDM index are characterized by a similar behavior.

Let us first examine the intraday dynamics of the cryptocurrency market portfolio approximated by the eigenportfolio from hourly data to see its periodic patterns.
Figure 3.3 shows the average hourly patterns in the cryptocurrency market returns. We observe, on average, positive returns at UTC hours 5:00, 8:00 and 20:00. These times correspond to the effects of Hang Seng and the closing of LSE and NYSE, respectively.

Figure 3.3: Average Hourly Eigenportfolio Return



This plot shows the average hourly returns of the hourly eigenportfolio returns by removing returns that have absolute values greater than two standard deviations. The vertical lines indicate the 25th and 75th percentiles of sample density.

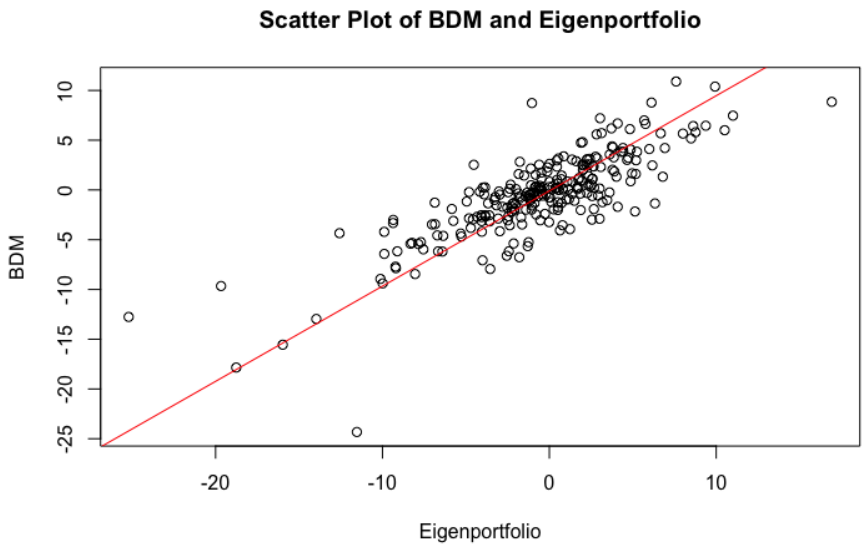
The intraday patterns of returns on the eigenportfolio are consistent with those revealed in the returns on native cryptocurrency and tokens in Sections 3.1.1 and 3.1.2 and illustrated in Figure 3.3, Section 3.1.2.

The daily eigenportfolio can be compared to the SPCBDM index, launched in July 2021. "The SP Cryptocurrency Broad Digital Market (BDM) Index is designed to track the performance of digital assets listed on recognized open digital exchanges that meet minimum liquidity and market capitalization criteria that are covered by (our) price provider Lukka" The index is meant to reflect a broad investable universe." [see, https://www.spglobal.com/spdji/en/indices/digital-assets/sp-cryptocurrency-broad- digital-market-index/#overview].

The index is part of an expansion of SP DJI's series of digital asset benchmarks, the SP Digital Market Indices. All these indices use pricing data from Lukka, a crypto software and data provider, to determine the eligibility universe and pricing of individual constituents.

The main difference between these indexes and our market portfolio is that the indexes are available only daily between Monday and Friday, while the computed eigenportfolio is available for all days and hours. Moreover, the index allocations are determined by Lukka, while our approach follows from the financial theory (Avellaneda et al., 2022). Nevertheless, we can select the returns on the market portfolio over the workdays to compare them to the SPCBDM. The linear regression results suggest that the two variables are collinear. The estimated intercept is not statistically significant (-0.1000, standard error 0.1827), and the estimated coefficient is 0.95674 with a standard error of 0.0427. The asymptotically valid confidence interval at level 0.95 contains the value 1. The scatter plot and the regression line are given in Figure 3.4.

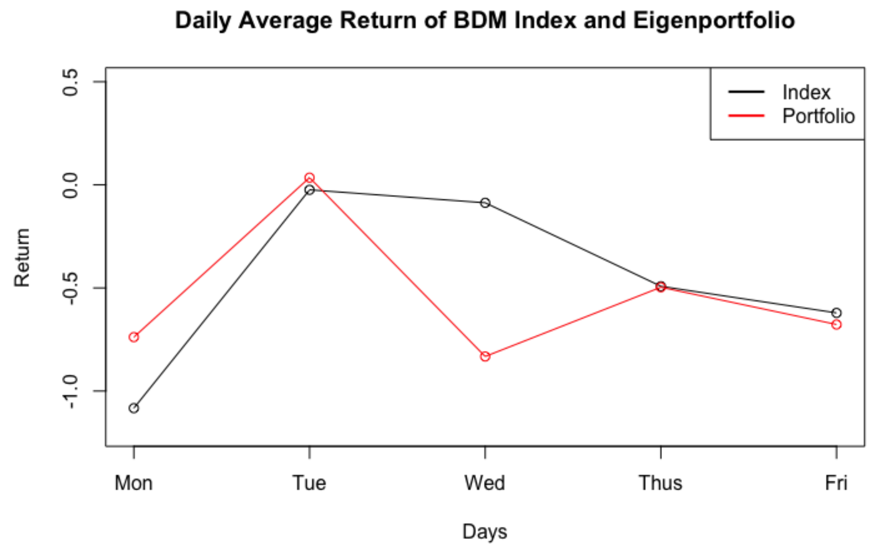
Figure 3.4: The Scatter plot of the SPCBDM Index and Eigenportfolio



The scatter plot reveals collinearity between daily returns on cryptocurrency eigenportfolio and the SPCBDM Index

Figure 3.5 compares the daily patterns of the SPCBDM index with the daily returns on the eigenportfolio.

Figure 3.5: The Average Return of the SPCBDM index and Eigenportfolio



The black line shows the average return of SPCBDM Index. The red line shows the average return on the eigenportfolio. The two lines overlap except on Wednesdays.

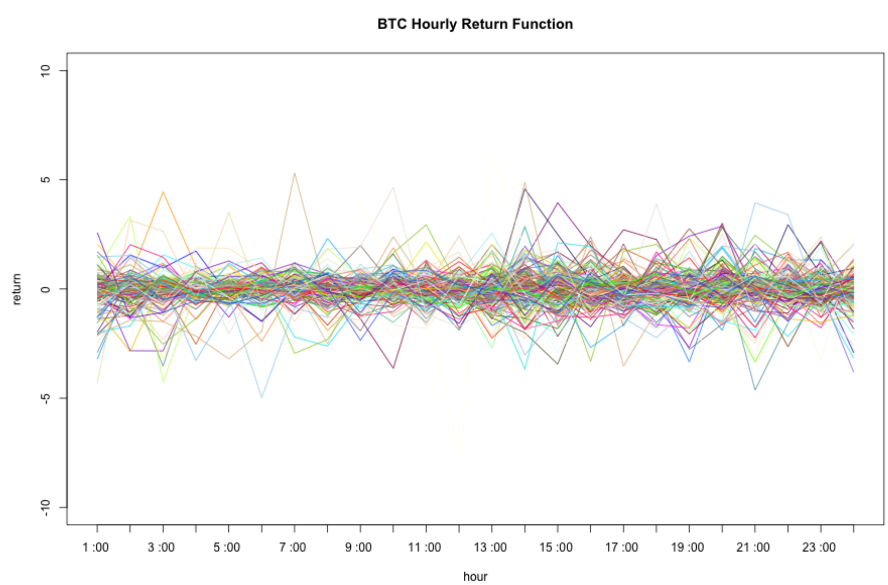
We find that both return series have similar daily patterns, except for Wednesdays when the returns on the index are higher than on the eigenportfolio due to extremely negative returns on the eigenportfolio on May 11, 2022 and November 09, 2022. Given these results, we conclude that the eigenportfolio provides a good approximation of the cryptocurrency market factor.

3.4 Functional Regression

In this section, we introduce an extension of the CAPM for cryptocurrency that accounts for periodicity by considering the relationship between the functions of returns over an entire day or week.

According to the CAPM, a contemporaneous relationship between the returns on individual coins and the cryptocurrency market portfolio would reflect the systemic risk in the cryptocurrency market. This Section estimates the relationships between the functions of daily and weekly returns on cryptocurrency and the returns on market factor to find out if such a relationship exists, while capturing the periodic intraday and intraweek patterns in systemic risk. If an affine contemporaneous relationship exists, then we could interpret its coefficients as the "betas" of a functional CAPM model and study their behavior across a day and week. In addition, we can address the endogeneity of cryptocurrency price dynamics, determined from a functional CAPM equilibrium. We consider the returns on BTC as the first example and compute the excess returns by using the US Three-Month Treasury Bond as the risk-free asset [5]. Figures 3.6 and 3.7 below present the 178 functions of returns on BTC over 24 hours and 178 functions of return on the cryptocurrency market factor over 72 hours, respectively.

Figure 3.6: The Hourly Return Function of BTC

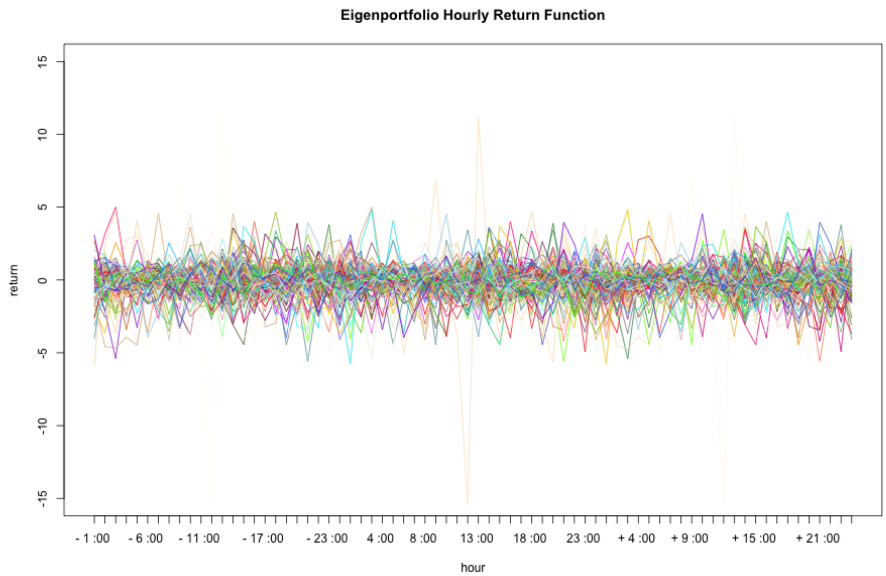


Each line represents the function of BTC hourly returns over one day. These functions are the observations on the dependent variable.

[5] The hourly excess returns are calculated as the log return less the three-month T-bill. The T-bill rates are divided by 365×24 to convert to an hourly rate. The data are downloaded from https://www.marketwatch.com/investing/bond/tmubmusd03m?countrycode=bx&mod=mw_quote_tab

Figure 3.6 presents the "dependent" variable functions of excess hourly returns on each of the 178 days in sample.

Figure 3.7: The Market Factor Hourly Return Functions



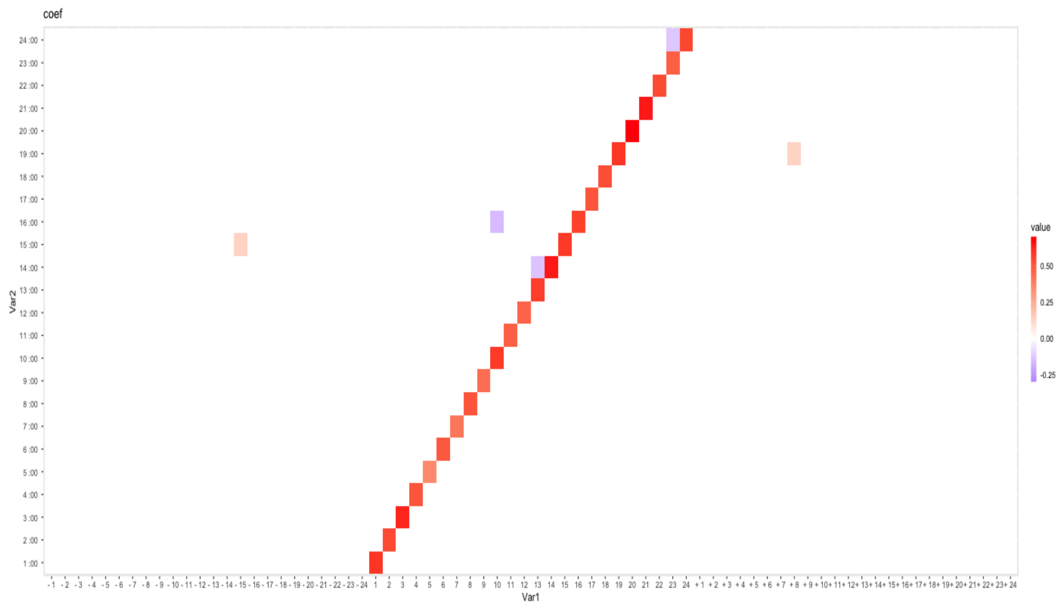
Each line represents the function of market factor hourly returns over three days. These functions are the observations on the regressor.

Figure 3.7 shows the "regressor" functions over the past, current and future days. The functional regression contains the lagged returns on the market factor in the regressor functions. The statistical significance of coefficients on these lagged market returns can be seen as evidence against the one-factor "static" CAPM.

3.4.1 Hourly Functional Regression

In the functional hourly returns model, the excess returns on a coin are regressed on the excess market return on the current, previous and future days to capture the lead and lag effects in the coefficients. Figure 3.8 presents the statistically significant components of the estimated integral operator Π of BTC returns. We observe an affine relation over the current day. This is consistent with the CAPM model, suggesting a linear relationship between the contemporaneous excess returns on an asset and the market. The dark red points on the line indicate the times of the day when the relationship is positive and high, whereas the lighter points indicate lower values of "betas". The difference in colors across the day suggests that the BTC "betas" vary depending on the time of the day and are higher at UTC hour 1:00 associated with the opening of Hang Seng, UTC hour 14 associated with the opening of NYSE and UTC hours 19, 20 and 21 associated with the closing of NYSE and the hours following the closure. The "betas" are lower at UTC hour 5 (mid-day of Hang-Seng) and hours 7 and 9 before and after the opening of LSE. (In Figure 3.8 the coefficients at hours 1, 3, 14, 19, 20, and 21 have relatively darker red color than at hours 5, 7, and 9.) In addition, we observe some lagged coefficients of current (excess) returns on Bitcoin on previous (excess) returns on the market portfolio. This occurs, for example, at UTC hour 15:00, which is one hour before LSE closes and after NYSE opens (Long et al., 2020).

Figure 3.8: The Heatmap of Hourly Functional Regression Coefficients for BTC



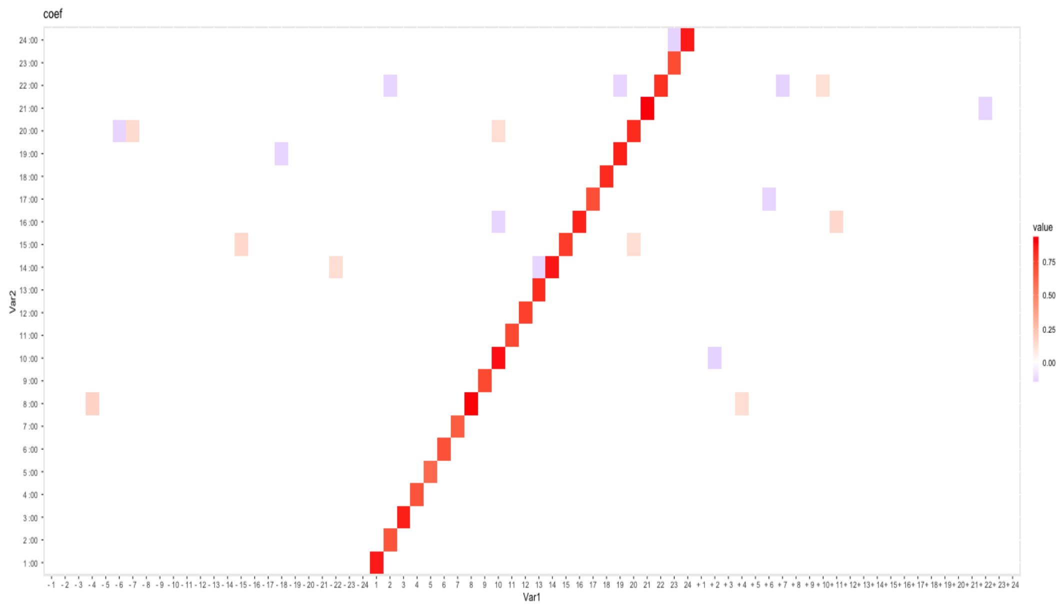
The red and light red areas indicate that the coefficients are positive or slightly positive. The light blue areas indicate that the coefficients are slightly negative. The time labels on the horizontal and vertical axes are UTC hours. The "-" and "+" on the horizontal axis indicate the previous day and the future day

To check the validity of the functional CAPM model, we estimate the intercepts of the functional regression. The estimated intercept values reported in Table A.5 of the Appendix are not statistically significant, except for UTC hour 19:00.

We conclude that Bitcoin returns at each hour seem to satisfy a one-factor CAPM model, except for hours 19:00 when the intercept is statistically significant and hours 14:00-16:00 and 0:00 when BTC returns depend not only on the current but also on the lagged market factor returns.

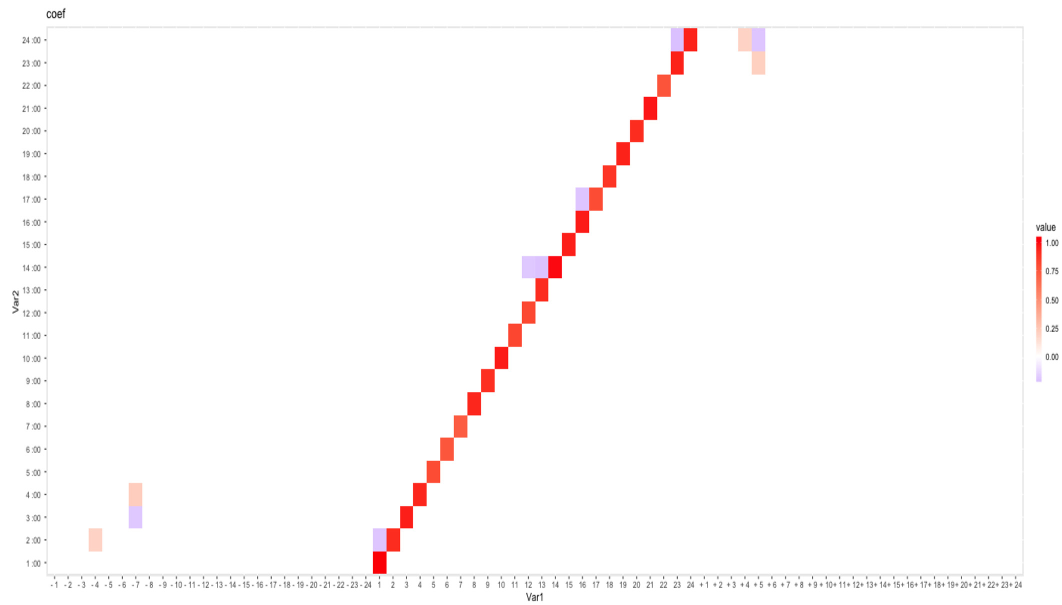
The affine relationship between the excess returns on individual cryptocurrency and market portfolio exists for most natives and tokens. It is illustrated in Figure 3.9 for ETH and Figure 3.10 for LINK.

Figure 3.9: The Heatmap of Hourly Functional Regression Coefficients for ETH



The red and light red areas indicate that the coefficients are positive or slightly positive. The light blue areas indicate that the coefficients are slightly negative. The time labels on the horizontal and vertical axes are UTC hours. The "-" and "+" on the horizontal axis indicate the previous day and the future day

Figure 3.10: The Heatmap of Hourly Functional Regression Coefficients for LINK



The red and light red areas indicate that the coefficients are positive or slightly positive. The light blue areas indicate that the coefficients are slightly negative. The time labels on the horizontal and vertical axes are UTC hours . The "-" and "+" on the horizontal axis indicate the previous day and the future day

We find that the "betas" vary slightly across the day in both Figures 3.9 and 3.10.

The ETH "betas" are higher at UTC hours 9 and 21, which is associated with the opening hour of LSE and the closing hour of NYSE. In the meanwhile, at UTC hours 4,5 and 7, we observe lower "betas" during the second-half of the trading day of Hang Seng.

The LINK "betas" show similar patterns and are high at UTC hours 1 and 14, which are associated with the opening hours of Hang Seng and NYSE. At UTC hours 6 and 7 we observe lower "betas" before the closing of Hang Seng.

In comparison to BTC, the returns on ETH seem more affected by the lagged and future market returns after hour 14:00, i.e. after the opening of NYSE. The effects of future market returns can be explained by the trades of derivatives on the ETH platform. The return on LINK resembles BTC more in that there are fewer times of the day when we observe the impact of lagged market returns, except at 14:00 (NYSE) and between hours 0:00 and 5:00 when Hang Seng is open. The negative values of coefficients on the lagged market returns marked in blue in Figures 3.9 and 3.10 suggest the presence of reversal effects and potentially multiple factors, in line with Shen et al. (2020).

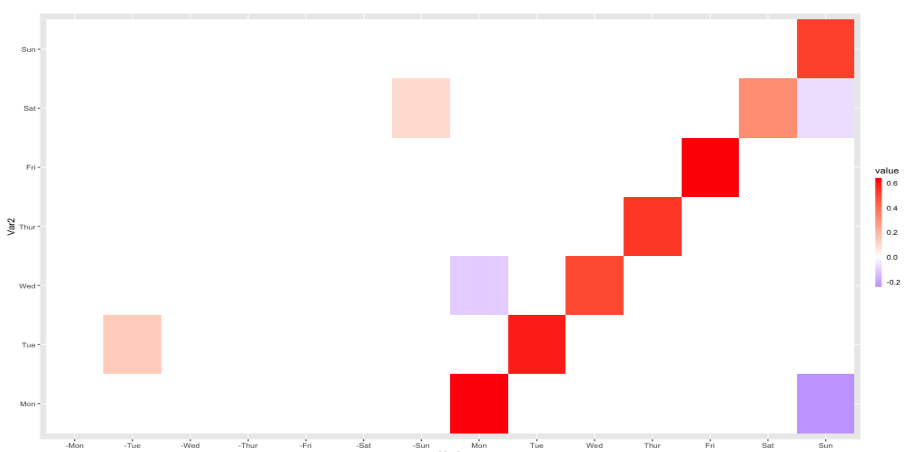
The affine relationship with the cryptocurrency market portfolio is not found in hourly returns on stablecoins, native SGB and token KNC that were found to be uncorrelated with other coins in Section 5 (see, Jasiak and Zhong (2024, Appendix B)).

3.4.2 Daily Functional Regression

The daily functional regressions provide the values of "betas" on each day of the week. We regress 52 functions of weekly excess returns on a coin on the functions of weekly returns on the market portfolio over the past and current week to capture the potential lag effects [6].

Figure 3.11 presents the estimator of the daily integral operator of the functional regression of excess BTC returns on the market portfolio.

Figure 3.11: The Heatmap of Daily Functional Regression Coefficients for BTC



The red and light red areas indicate that the coefficients are positive or slightly positive. The light blue areas indicate that the coefficients are slightly negative

[6] The future week is not included because of the limited number of functions available for the daily analysis.

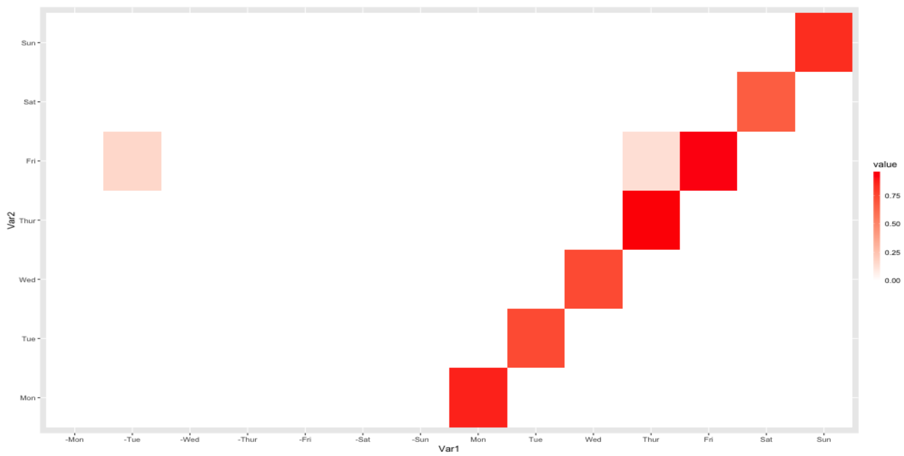
We observe an affine contemporaneous relationship between the current excess returns on BTC and the cryptocurrency market portfolio. The "betas" are positive and high on all days, except for Saturdays when they are slightly weaker. We do not observe any effects of lagged market returns on Thursdays, Fridays and Sundays. There are some significant lagged coefficients, indicating, for example, the positive dependence of BTC Saturday returns on the market returns from the previous Sunday and the negative dependence of BTC Monday returns on the current week's Sunday market returns.

To check the validity of the daily functional CAPM for BTC, we compute the intercepts and report them in Table A.6 of the Appendix. All the intercept terms are not statistically significant.

Given these results, we conclude that the daily BTC returns are consistent with the one-factor CAPM model on Thursdays, Fridays and Sundays in the sense that we do not find statistically significant effects of past market returns.

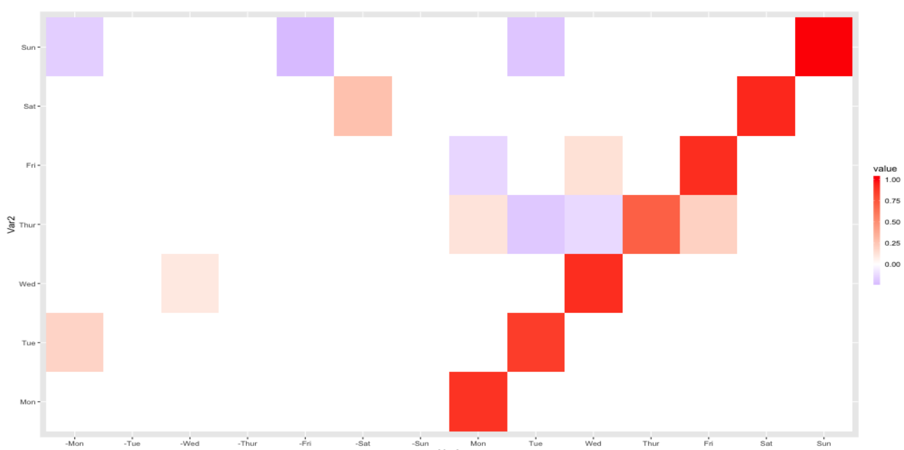
We find affine functional relations between the market factor returns and the daily returns on most natives and tokens. Figure 3.12 illustrates the functional "betas" for ETH and Figure 3.13 for LINK. The coefficients of excess daily returns on the market portfolio are lower on Saturdays in the BTC and ETH functional regressions. In LINK, the lowest coefficient is associated with Thursday returns.

Figure 3.12: The Heatmap of Daily Functional Regression Coefficients for ETH



The red and light red areas indicate that the coefficients are positive or slightly positive.

Figure 3.13: The Heatmap of Daily Functional Regression Coefficients for LINK



The red and light red areas indicate that the coefficients are positive or slightly positive. The light blue areas indicate that the coefficients are slightly negative

There are also some lagged effects revealed in Figures 3.12 and 3.13. The Friday returns on ETH are positively related to the market factor returns on the previous day, i.e. Thursday and the past week's Tuesday. The LINK's returns on Thursdays and Fridays show dependence on market factor returns from the previous days of the same week with either positive or negative coefficients. On Sundays, Link has negative coefficients on the current week's Thursday market returns and the past week's Friday and Monday market returns. This can be considered as evidence of reversal effects or additional market factors suggested by Shen et al. (2020)

The affine relationship is not found in all coins and on all days of the week. For example, SBG has only one statistically significant coefficient on Wednesday. KNC has no statistically significant coefficients at all.

The daily returns on stablecoins show no relation with the daily market returns, which is consistent with the results from the hourly regression.

3.4.3 Endogeneity of Price Dynamics

Let us now reconcile the empirical evidence provided in Sections 6.1 and 6.2 with the CAPM equilibrium. The intraday functional regression of excess cryptocurrency returns $y_i(s)$ on excess cryptocurrency returns $x_i(s)$ introduced in Section 4.4 is:

$$y_i(s) = \int_{-24}^{48} \Pi(s,t) x_i(t) dt + u_i(s),$$

where i denotes the day and s, t are the hours of the day. This model can be summarized for $i = 1, ..., n$ as:

$$Y_i = \Pi X_i + U_i, \quad i = 1, ..., n$$

where Y_i and X_i are in separable Hilbert spaces of square integrable functions $\mathcal{E} = L^2[0, 24]$, $\mathcal{F} = L^2[-24, 48]$, respectively, U_i is a zero mean random element of $\mathcal{E}$, and Π is a non-random Hilbert-Schmidt operator from $\mathcal{F}$ to $\mathcal{E}$. The discrete analogous model, written for simplicity over the past 24 hours only, is:

$$y_i(s) = \sum_{t=-24}^{-1} \pi(s,t)x_i(t) + \pi(s,s)x_i(s) + u_i(s).$$

This is a model of coin excess returns with a past dependent intercept $\sum_{t=-24}^{-1} \pi(s,t)x_i(t)$ and beta $\pi(s,s)$ that depends on the hour of the day (or day of the week in daily data). It can be compared with the CAPM model written as a conditional affine regression of excess returns on a coin on the excess return on the cryptocurrency market:

$$r_t = \alpha_{t-1} + \beta_{t-1} r_{m,t} + e_t$$

where $r_t, r_{m,t}$ are the excess returns on an asset and the market portfolio, respectively, $E_{t-1}(e_t) = 0, Cov_{t-1}(e_t, r_{m,t}) = 0$ and the coefficients $\alpha_{t-1}, \beta_{t-1}$ are functions of the past. At equilibrium $\alpha_{t-1} = 0$ (Gourieroux and Jasiak, 2001). Let us define an analogous equilibrium condition at time s under the assumption of static and exogenous asset supply and dynamic excess returns. Then, for the day and time i, s, we get:

$$y_i(s) = \alpha_i(\underline{y_i}(s)) + \beta_{s,s} x_i(s) + u_i(s)$$

where the excess market returns and the excess returns on a coin over 24 hours prior to s are denoted by $\underline{x_i}(s)$ and $\underline{y_i}(s)$, respectively and $\underline{x_i}(s) \subset \underline{y_i}(s)$. If the equilibrium holds at a given hour of the day (or day of the week for daily data), then the CAPM model implies that the intercept term is zero:

$$\alpha_i(\underline{y_i}(s)) = 0$$

The above condition is satisfied, specifically, when all the coefficients $\pi(s,t)$ on the past returns over the 24 hours prior to s (or 7 days prior in daily data) and the constant term are zero. Then, the cryptocurrency price dynamics are endogenously determined from the CAPM equilibrium at a given hour of the day, or day of the week.

Our empirical evidence shows that the CAPM equilibrium is not always satisfied, and selected past market return coefficients can be statistically significant. Specifically, the equilibrium condition in hourly data is observed more often in Bitcoin and Link than in Ether. In daily data, the equilibrium seems to be satisfied more often in Ether than in Bitcoin and Link. However, it is possible that if the analysis was conditioned on the cryptocurrency market only and any stock market effects were excluded, the CAPM condition would hold. We observe that most non-zero intercepts in hourly data appear periodically because of the effects of the openings and closures of the major stock markets.

In our model for hourly data, we have also included the cryptocurrency market returns over the 24 hours posterior to time s. This can be justified by the rational expectations of investors who observe periodic patterns and anticipate the effects of the stock market over the remaining hours of the day.

The intraday periodicity determined by the opening and closures of the major stock markets indicates that the cryptocurrency and stock markets are inter-related at some hours. This suggests that there exists an overall financial market, and a segmentation of that overall financial market into separate cryptocurrency and stock markets is not always plausible.

In this context, we can also address the endogeneity of the cryptocurrency market return variable. In the functional regressions of Sections 6.1 and 6.2, the first principal component is used as the proxy of the cryptocurrency market portfolio. We know from Section 5.2 that the value of the first principal component portfolio is collinear with the daily observed SPCBDM Index value, and their intraweek periodic patterns are similar. However, they are both cryptocurrency-based proxies of a latent true market return.

According to the Roll's critique and recent financial literature, the use of a proxy in a CAPM model leads to misspecification that manifests itself through under-estimated "betas" and non-zero intercepts (Prono, 2015; Simmet and Pohlmeier, 2020). The source of misspecification is the latency of the market return (Prono, 2015). Simmet and Pohlmeier (2020) point out that a proxy is necessarily based on only a subset of all assets and the weights associated with each of them may differ from the true weights. In our case, the market proxy is obtained from the PCA analysis which divides the cryptocurrencies into orthogonal subsets, so that the coefficients remain unchanged if the model includes only a subset of them.

Prono (2015) argues that "the proxy return is endogenous...by nature of the measurement error associated with it" (Prono, 2015). Then, if the market proxy p in $\mathcal{F}$ is such that:

$$\forall i, s: \quad p_i(s) = x_i(s) + \zeta_i(s)$$

where the the measurement error $\xi_i(s)$ has mean zero, then, a model with the proxy p used instead of the true market excess return x suffers from an econometric endogeneity problem and the estimator of beta is downward biased leading to a non-zero intercept.

In Sections 6.1 and 6.2, we observed that non-zero intercepts arise occasionally. Therefore, on the cryptocurrency market, the eigenportfolio (excess) return can be arbitrarily close to the true market return at selected times, and the discrepancy between the market proxy and the true market (excess) return could be triggered by the periodic patterns described in Section 3.

3.5 Conclusion

The correlation analysis shows a strong contemporaneous dependence between the returns on native cryptocurrencies and tokens. The stablecoin returns are uncorrelated with other cryptocurrencies and are also uncorrelated with one another. The PCA applied to daily returns suggests the presence of one factor in the cryptocurrency market to which the stablecoins do not contribute. Therefore, they are immune to the systemic risk associated with the cryptocurrency market. The functional regression defines an affine relation of cryptocurrency excess returns with the excess returns on the market portfolio. The "betas" are weaker or stronger, depending on the time of the day and day of the week. This suggests that the exposure to systemic risk displays periodic intraday and intraweek patterns. The affine relation does not exist for the stablecoins and coins, which are uncorrelated with the cryptocurrency market.

More research is needed regarding the stability of periodic patterns revealed in this study. In addition, it would be interesting to address the complexity of systemic risk on the cryptocurrency market, and the stock market effects evidenced by the functional CAPM. An advanced analysis of the theoretical foundations of the functional CAPM and the inference on functional betas is left for further research.

Chapter 4

A New Approach for Functional Principal Component Analysis (FPCA) in Bitcoin Return Forecasting

Coauthorship Statement

This chapter is based on joint work titled "Intraday Functional PCA Forecasting of Cryptocurrency Returns" with Dr. Joann Jasiak. All authors contributed equally to the conceptualization, data preparation, empirical modeling, analysis, and writing of the manuscript.

A version of this chapter is currently accepted and published at `Journal of Forecasting`.

4.1 Introduction

Bitcoin, the world's first cryptocurrency introduced in 2008, has become one of the most important digital assets worldwide. Its decentralized, peer-to-peer network, underpinned by blockchain technology, ensures transparency, security, and independence from traditional banking systems. Like other cryptocurrencies, Bitcoin is traded 24 hours a day, 7 days a week. Recently, Bitcoin prices have increased considerably and its returns remained highly volatile. The high return volatility and speculative nature of Bitcoin present both opportunities and challenges for investors. Bitcoin's dynamics are influenced by a complex interplay of factors, including market sentiment, regulatory developments, technological advancements, macroeconomic trends, and geopolitical events. These factors introduce substantial uncertainty, making Bitcoin forecasting particularly challenging.

Several studies have employed various methods to forecast Bitcoin returns and volatility. Katsiampa (2017) applied traditional time series models, such as GARCH, to estimate Bitcoin's volatility. Bouri et al. (2017) explored the return-volatility relationship in the Bitcoin market before and after 2013 using an asymmetric GARCH model. Gradojevic et al. (2023) applied linear time series models and nonlinear predictors, such as feedforward deep artificial neural network(FF-D-ANN), support vector machine(SVM) and random forest(RF) to forecast daily and hourly Bitcoin returns from exogenous variables. Wen et al. (2022) documented strong intraday return predictability in Bitcoin, Ethereum, and Ripple, identifying both momentum and reversal effects. Yarovaya and Zięba (2022) examined the volume–return linkages, using five-minute data for 30 cryptocurrencies, and found bidirectional causality between trading volume and returns. Ahmed and Al Mafrachi (2021) showed that high-frequency realized higher-order moments contain predictive power for short-term crypto returns. Bouri et al. (2021) used functional data analysis to forecast Bitcoin's intraday prices from cumulative intraday return (CIDR) curves, demonstrating profitable intraday trading strategies. Machine-learning studies such as Gyamerah and Cai (2021) and Rodrigues and Machado (2025) highlight the improved short-horizon accuracy of hybrid or recurrent neural architectures (VMD-SVR, GRU) over ARIMA-type models in minute-level Bitcoin price forecasts.

Since Bitcoin is traded continuously, its log-returns[1] on a given day can be viewed as a function of time measured in minutes or seconds of UTC[2] [see **?** for Bitcoin returns, and e.g. Aue et al. (2017), Kokoszka et al. (2017) for applications to intraday returns on stocks, stock market indexes, including S&P500 and ETF's]. Then, the daily return functions can be modeled as a functional time series and predicted, using the Functional Autoregressive process or order 1 (FAR(1)). Aue et al. (2015) show that this approach is asymptotically equivalent to predicting from the Karhunen-Loeve expansion of a functional time series and Vector Autoregressive of order 1 (VAR(1)) model of eigenscores obtained from a Functional Principal Component Analysis (FPCA) of the covariance operator. The advantage of the latter approach is that it can easily be adapted to the return dynamics characterized by conditional heteroscedasticity, i.e., time-varying volatility. The extension of the FPCA-based forecasting method to conditionally heteroscedastic data to accommodate time-varying volatility in financial returns is the main objective of this paper.

Our first contribution aims at improving interval forecasting of daily return functions and exploits serial correlation of eigenscores and squared eigenscores. The novelty of our approach is that by taking into consideration the conditional heteroscedasticity of eigenscores, we produce forecast intervals that are more accurate than those obtained when the conditional heteroscedasticity is disregarded. Since a functional dynamic joint model for serially correlated and conditionally heteroscedastic functional time series is not available in the literature, we introduce the Karhunen-Loeve (KL) Dynamic Factor model by imposing simplifying assumptions on the Karhunen-Loeve series expansion. This allows us to bridge the conditionally heteroscedastic discrete-time models with their functional counterparts. The KL Conditionally Heteroscedastic Dynamic Factor model offers a convenient framework for the analysis of conditionally heteroscedastic functional time series, such as returns on Bitcoin or other cryptocurrency, or equity stocks. It reveals the daily factors of volatility with intraday functional coefficients, offering a new insight into daily volatility and risk

[1]computed as the first differences of log-prices and referred to as returns henceforth

[2]Coordinated Universal Time

premia. The intraday functional factor loadings accommodate the shapes of return curves (Litterman and Scheinkman, 1991; Kokoszka et al., 2014) and illustrate the variation of risk exposure. In this context, the KL Conditionally Heteroscedastic Dynamic Factor model is comparable to the ARCH-M model of Engle et al. (1987) for discrete-time return data.

Our second contribution is a rolling FPCA algorithm that allows for intraday point forecasting of incomplete daily functions of Bitcoin returns. We observe that Bitcoin and many other financial assets are traded 24/7, and the delimitation of the day between hours 0:00 and 24:00 of UTC is purely conventional. This motivates us to consider the FPCA of partially overlapping daily return functions that start at subsequent time points to forecast a future segment of a function of current intraday returns. The proposed algorithm is simple and can be combined with machine learning methods (SVM, RF, NN) for improved performance.

Accurate financial forecasting is essential for investors as it enables informed decision-making, effective risk management, better financial planning, and gaining credibility with stakeholders like investors and lenders. It helps businesses anticipate future financial performance based on an analysis of historical data, market trends, and economic indicators. Makridakis et al. (2009) argue that improved forecast accuracy strengthens business decision-making under uncertainty, promoting investment confidence and risk-adjusted returns. Similarly, Hobbs et al. (1998) show that forecast improvements translate directly into cost savings and operational efficiency. Accurate forecasting of intra-day Bitcoin returns is important for day traders and market professionals to capitalize on volatility, manage risk, and exploit short-term profit opportunities in the 24/7 market. The high volatility of Bitcoin, which can fluctuate significantly within hours, means that intraday forecasting is a critical tool for making informed trading decisions and developing profitable strategies. The advantage of the proposed functional approach compared to the conventional forecasting of discrete-time returns is that the function can be evaluated at any time in a day and provide forecasts at any point of time during the day, rather than at a fixed sampling frequency.

In our empirical study, we examine a sample of daily functions of 15-minute and hourly returns. We model and forecast the return functions and evaluate the performance of our forecasting methods. Our method of forecasting daily return functions from the KL Dynamic Factor model produces narrower and more accurate forecast intervals based on predicted time-varying return volatility than those introduced in Aue et al. (2015). Our proposed algorithm for intraday forecasting of incomplete return functions based on a rolling FPCA is shown to compare favorably with the machine learning methods applied by Gradojevic et al. (2023) to Bitcoin returns in terms of forecast accuracy and sign at fixed future points of time.

In the literature, the Functional Principal Component Analysis (FPCA) has been used for forecasting first by Aguilera et al. (1997). Their method is applicable to forecasting long segments of future functions so that the functional time interval can be divided into the associated "past", i.e., t=1,..., T, and "future", i.e., t=T+1,..., T+K, say. The FPCA is then applied separately to the divided functional data relating to the "past" and "future" times, yielding past" and "future" eigenscores and eigenfunctions independently. The forecast future values of the function are estimated from a linear combination of the estimated "future" eigenscores and "past" eigenfunctions plus the estimated future mean. Our intraday FPCA forecasting is in part inspired by this method and allows for forecasting in real time, with the "future" potentially reduced to a neighborhood of one point in time.

Aue et al. (2015) proposed an approach to functional forecasting of complete future functions using the FPCA combined with a Vector Autoregressive (VAR(1)) model of eigenscores. This method was applied to environmental data of pollution concentrations that follow a stationary (non-trending) process. Shang (2020) used a similar approach, except that they used univariate autoregressive-moving average ARMA(p,q) time series models to forecast each eigenscore process independently in a study of Japanese age-specific mortality rates. Shang and Kearney (2022) applied independently the univariate ARMA(p,q) time series models to forecast the eigenscores in an empirical study of implied volatility of foreign exchange rates. Bouri et al. (2021) used the AR(1) model of eigenscores to forecast prices from the cumulative return functions (CIDR) of Bitcoin and showed its improved performance compared to the eigenscores predicted from exponential smoothing.

Our functional return forecasts improve upon the methods of Aue et al. (2015), Shang (2020) by adjusting them to the dynamics of financial returns. In particular, we observe that the eigenscores of returns are characterized by serial correlation at short lags, which can be parsimoniously modeled by the Autoregressive AR(1) process rather than an ARMA(p,q), in line with Bouri et al. (2021). Moreover, we evidence serial correlation in the squared eigenscores, i.e. we show that they are conditionally heteroscedastic and characterized by the ARCH effect. This characteristic is important, and by accounting for the conditional heteroscedasticity of eigenscores, we produce improved pointwise interval forecasts of return functions. In addition, to account for lagged cross-correlations between the eigenscores and their squares, we improve the approach of Aue et al. (2015) by replacing the VAR(1) forecast of eigenscores, by the VAR(1)-sBEKK(1,1) model that accommodates serial correlation, lagged cross-correlation, and conditional heteroscedasticity of eigenscores in a multivariate setup. The proposed methods are applicable both to the returns and cumulative returns (CIDR). The focus on returns instead of CIDR ensures the financial interpretation of the KL Dynamic Factor model as a time-varying risk premium model. This, along with evidencing the conditional heteroscedasticity of eigenscores and their lagged cross-correlations, distinguishes our approach from Bouri et al. (2021) and Wang et al. (2025).

The paper is organized as follows: Section 2 describes the FPCA and KL dynamic models, introduces the new forecast method of the forecast interval and compares it with the existing FPCA forecasting models. Section 3 describes the functional data on Bitcoin and presents the results of our forecasting method for daily return functions. Section 4 introduces a new method of forecasting the returns on Bitcoin from a rolling FPCA when the function of observations is incomplete, and compares it with other Bitcoin return forecast models. Section 5 concludes the paper. Additional results are provided in Appendices A, B, C, D.

4.2 Methods

This section reviews the Functional Principal Component Analysis (FPCA), introduces the new approach to functional times series forecasting of conditionally heteroscedastic returns, and compares it with the existing methods.

In the literature [see, e.g Ramsay and Silverman (2005b)] functional data are usually modeled as "noise-corrupted observations on a set of trajectories" that are assumed to be realizations of a smooth random function of time $X(t)$, with unknown mean shape $\mu(t)$ and covariance $c(t,s) = cov(X(t), X(s))$ (Liu et al., 2017).

Let $\{X_i(t)\}_{i\in\mathbb{Z}}$ be a stationary functional time series, where $t \in I = [a,b]$ is a continuum, and $i = 1, 2, ...$ denote the realization. Each realization can be written as a function $X_i(t,\omega)$ in both time $t \in I$ and $\omega \in \Omega$, an element of probability space. For a fixed ω, and for each i, $X_i(t)$ is in the Hilbert space $H = l^2(I)$, equipped with the inner product $\langle x, y\rangle_{l^2} = \int_I x(t)y(t)dt$ of square-integrable real functions on a support $I \subset \mathbb{R}$ and the norm $||x||^2_{l^2} = \int_I x^2(t)dt$ (Loeve, 1978; Shang, 2020). At a given t and for each i, $X_i(t,\omega)$ is a stochastic variable defined on a common probability space with finite second-order moment, i.e. any $X_i(t)$ is such that $EX_i(t)^2 < \infty\ \forall i, t$, where $E(X_i(t)^2) = \int_\Omega X_i^2(t,\omega)dP(\omega)$. The space $L^2(\Omega, \mathcal{A}, \mathcal{P})$ of variables with finite second-order moments is another Hilbert space equipped with the inner product $< X, Y >_{L^2} = \int_\Omega X(\omega)Y(\omega)dP(\omega)$.

4.2.1 The Functional Principal Component Analysis (FPCA)

The FPCA is based on the Karhunen-Loève expansion, which is a spectral decomposition of the covariance operator of the functional data. The Karhunen-Loève expansion can be written for each stochastic function $X_i = \{X_i(t),\, t \in [a,b]\}$ that is zero-mean, or demeaned, i.e. replaced by $X_i(t) - \mu(t)$, and where $\mu(t) = EX_i(t)$ is independent of index i by the stationarity assumption.

The covariance function $c(t, s)$ of a stochastic function $X(t)$ is:

$$c(t, s) = E[X(t) - \mu(t)][X(s) - \mu(s)], \; \forall t, s \in I, \tag{4.2.1}$$

where $\mu(t)$ is the mean function. The covariance function $c(t, s)$ allows the covariance operator of X denoted by K to be defined as:

$$K(\xi)(s) = \int_I c(t, s)\xi(t)dt. \tag{4.2.2}$$

This operator is a function from $l^2(I)$ onto $l^2(I)$ that associates the element $K(\xi) \in l^2(I)$ with any element ξ of $l^2(I)$. This operator is continuous, symmetric, positive, i.e. $< \xi, K(\xi) >_{l^2(I)} \geq 0$, $\forall \xi \in l^2(I)$. It follows from Mercer's and Karhunen-Loève Theorems that there is an orthonormal sequence (ξ_j) of functions in $l^2(I)$ and a non-increasing sequence of positive numbers (λ_j) such that the operator K admits a spectral decomposition:

$$c(t, s) = \sum_{j=1}^{\infty} \lambda_j \xi_j(t) \xi_j(s), \qquad t, s \in I \tag{4.2.3}$$

where $\lambda_j, j = 1, 2, ...$ are the eigenvalues in strictly decreasing order and $\xi_j(s), j = 1, 2, ..$ are the normalized eigenfunctions, so that $K(\xi_j) = \lambda_j \xi_j$ with $< \xi_j, \xi_k >_{l^2} \equiv \delta_{j,k} = 1$, if $j = k$, and 0, otherwise. Moreover, ξ_j are continuous functions if the associated λ_j are strictly positive and the equality corresponds to both a convergence in L^2 for the Karhunen-Loève Theorem and a uniform convergence in Mercer's Theorem. Then, each stochastic function $X_i(t, \omega)$, $i = 1, 2, ...$ can be expressed as a linear combination of these basis functions.

Karhunen-Loève expansion: A stochastic function $X(t, \omega)$ can be expressed as a linear combination of orthogonal basis functions ξ_j determined by the covariance function of the process:

$$X(t, \omega) = \mu(t) + \sum_{j=1}^{\infty} \beta_j(\omega) \xi_j(t), \tag{4.2.4}$$

or, by omitting ω for ease of exposition, as:

$$X(t) = \mu(t) + \sum_{j=1}^{\infty} \beta_j \xi_j(t), \tag{4.2.5}$$

where β_j are eigenscores obtained from a projection in $l^2(I)$ of $[X(t)-\mu(t)]$ in the direction of the jth eigenfunction ξ_j, i.e. $\beta_j =< [X(t)-\mu(t)], \xi_j(t) >_{l^2(I)}$. The eigenscores $\beta_j,\ j = 1, 2, ..J$ are pairwise uncorrelated random variables with zero mean and variance λ_j (Benko et al., 2009). Each of them can be interpreted as the contribution of $\xi_j(t)$ to $X(t) - \mu(t)$. The functions ξ_j are continuous real-valued deterministic functions on $I = [a, b]$ that are pairwise orthogonal in $l^2(I)$, i.e., in the time domain. If $X(t)$ is Gaussian, then the random variables β_j are also Gaussian, uncorrelated, and then stochastically independent. Alternatively, we can write:

$$X(t, \omega) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\lambda_j} \beta_j^*(\omega) \xi_j(t), \tag{4.2.6}$$

to see that $\beta_j^*(\omega)$ are orthonormal in $L^2(\Omega, \mathcal{A}, \mathcal{P})$, and ξ_j are orthonormal in $l^2(I)$. It is therefore sometimes said that the Karhunen-Loève expansion is bi-orthogonal (bi-orthonormal).

4.2.2 The KL Dynamic Factor Model

We consider the stationary functional time series indexed by $i = 1, ..., N$ and the Karhunen-Loève (KL) expansion of $X_i(t)$ for each i:

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\lambda_j} \beta_{i,j}^* \xi_j(t), \tag{4.2.7}$$

where the stochastic eigenscores $\beta_{i,j}^*$ are zero-mean with unit variance, uncorrelated for different j, stationary in i, but can be correlated for different i. The above expansion is a representation of a stationary functional process (stationary in i) and imposes no other restriction, except for some weak regularity conditions. It can be written as:

$$X_i(t) = \mu(t) + \sum_{j=1}^{J} \sqrt{\lambda_j} \beta_{i,j}^* \xi_j(t) + e_i(t), \tag{4.2.8}$$

where $e_i(t) = \sum_{j=J+1}^{\infty} \sqrt{\lambda_j} \beta_{i,j}^* \xi_j(t)$. The truncation error function, denoted by $e_i(t)$, is stationary, with mean zero and finite variance. The tuning parameter J represents the number of eigenfunctions that are preserved in the approximation, resulting in a dimension reduction (Shang and Kearney, 2022).

The representation (4.2.8) can be used to define a structural model for functional time series by specifying the dynamics of the eigenscores, either $\beta_{i,j}^*$, or $\beta_{i,j} = \sqrt{\lambda_j} \beta_{i,j}^*$. More precisely, we make the following assumption:

Assumption 1:

i) There exists a true value J_0 of J.

ii) The eigenscore series $\beta_{i|J_0} = (\beta_{i,1}, ..., \beta_{i,J_0})$ and $\beta_{i|\bar{J}_0} = (\beta_{i,J_0+1}, \beta_{i,J_0+2}, ...)$ are independent.

iii) The conditional distribution of $\beta_{i|J_0} = (\beta_{i,1}, ..., \beta_{i,J_0})$ given $\underline{\beta_{i-1|J_0}} = (\beta_{i-1|J_0}, \beta_{i-2|J_0}, ...$ has a parametric specification with parameter θ.

iv) The remaining $\beta_{i|\bar{J}_0} = (\beta_{i,j|j>J_0})$ are independent white noises.

Definition 1: The KL dynamic factor model satisfies the decomposition (4.2.8) and Assumption 1.

It is easily checked that the KL expansion appears as a dynamic factor model with J_0 dynamic factors $\beta_{i|J_0}, i = 1, ..., J_0$ and an independent and identically distributed (i.i.d.) functional noise $e_i(t)$. The noise is partly degenerate since, by construction, we have:

$$< e_i, \xi_j >_{l^2(I)} = 0, \ \forall j = 1, ..., J_0, \tag{4.2.9}$$

and the support of its distribution is in the vector space orthogonal to the space of $\xi_j, j = 1, ..., J_0$ in $l^2(I)$.

Example 1: KL Linear Dynamic Factor Model

For ease of exposition, we consider a Gaussian framework with stationary linear dynamics of order 1. Then, the specification is:

$$X_i(t) = \mu(t) + \sum_{j=1}^{J_0} \beta_{i,j}\xi_j(t) + e_i(t), \tag{4.2.10}$$

where

$$\beta_{i|J_0} = \Phi\beta_{i-1|J_0} + \epsilon_i, \tag{4.2.11}$$

where $\beta_{i|J_0}$ is a vector of length J_0 containing $\beta_{i,j}, j = 1, ..., J_0$, where the eigenscores β_j are random variables with zero mean, the autoregressive matrix Φ of dimension $J_0 \times J_0$ has eigenvalues inside the unit circle, $(\epsilon_i), i = 1, 2, ..$ is an independent and identically distributed (i.i.d.) white noise of dimension J_0, $\epsilon_i \sim N(0, \Sigma), \forall i$, assumed to be independent of the functional noise $\{e_i(t), t \in [a, b]\}, i = 1, ..., N$. The functional noise has a functional Gaussian distribution of infinite dimension, denoted by $N(0, \sigma^2\Omega_e)$, where $\Omega_e = Id - \sum_{j=1}^{J_0} \xi_j\xi_j'$, and Id denotes the Identity operator on $l_2(I)$ and $\xi_j'\Omega_e\xi_j = 0, j = 1, ..., J_0$[3]. The model (4.2.10)-(4.2.11) depends on the parameters of different types. Some of them, such as Φ, Σ, σ^2, $\lambda_j, j = 1, ..., J_0$ are scalar or multidimensional parameters, while $\mu, \xi_j, j = 1, ..., J_0$, are functional parameters. Any daily function has a mean curve $\mu(t)$, which in applied research is removed from the data. A priori, there are no assumptions on the stationarity in t other than square integrability, implying finite first two moments.

The above model can be viewed in another way. It is equivalent to consider the functional series $X_i(t)$, $t \in I$, or the countable set of series $\beta_{i,N} = (\beta_{i,1}, ..., \beta_{i,j}, ...)$ that are related by a "one-to-one" linear relationship. Therefore, it is equivalent to write a Gaussian autoregressive model for the X's or for the β's. In the framework of this example, we have introduced a constrained Gaussian VAR(1) specification for the β's. Therefore, a reduced-rank, con-

[3]As pointed out later in Section 4-2.2.2, $\sum_{j=J_0+1}^{\infty} \xi_j\xi_j'$ and $\sum_{j=1}^{J_0} = \xi_j\xi_j'$ are orthonormal projectors.

strained Functional Autoregressive Gaussian FAR(1) is the associated representation of the functional process. The relationship between these models can be explained as follows. The FAR(1) model is:

$$X_i - \mu = \Psi(X_{i-1} - \mu) + u_i, \tag{4.2.12}$$

where the time t index is omitted for ease of exposition. Under the approach of Bosq (2000) the errors $(u_i, i \in \mathbb{Z})$ are centered (i.i.d.) innovations in $l_2(I)$ satisfying $E||u_i||^4 = E[\int u_i^2(t)dt]^2 < \infty$ and Ψ is a bounded linear operator satisfying the condition $\int\int \psi^2(s,t)dsdt < 1$ that ensures that the above equation has a strictly stationary and causal solution such that u_i is independent of $X_{i-1}, X_{i-2}, ...$ Suppose that $\mu = 0$ and J_0 is given by Assumption 1 i). Then, it follows from Theorem 3.1 of Aue et al. (2015)[4] that for $J = J_0$ the process can be written as

$$\begin{aligned} < X_i, \hat{\xi}_j > &= < \Psi(X_{i-1}), \xi_j > + < u_i, \xi_j > \\ &= \sum_{j'=1}^{\infty} < X_{i-1}, \xi_{j'} >< \Psi(\xi_{j'}), \xi_j > + < u_i, \xi_j > \\ &= \sum_{j'=1}^{J_0} < X_{i-1}, \xi_{j'} >< \Psi(\xi_{j'}), \xi_j > + \epsilon_{i,j} \end{aligned} \tag{4.2.13}$$

where $\epsilon_{i,j} = d_{i,j} + < u_i, \xi_j >$ with the remainder terms $d_{i,j} = \sum_{j'=J_0+1}^{\infty} < Y_{i-1}, \xi_{j'} >< \Psi(\hat{\xi}_{j'}), \hat{\xi}_j >$ being independent white noises by Assumption 1 iv). Next, the empirical counterpart of the above inner product $\beta_i =< X_i, \hat{\xi}_j >$ is written as follows. Let $\boldsymbol{\beta}_i = (\beta_{i,1}, \beta_{i,2}, ..., \beta_{i,J_0})'$ be a $(J_0 \times 1)$ vector of eigenscores on day i, to get a VAR(1) model:

$$\boldsymbol{\beta}_i = \mathbf{\Pi}_1 \boldsymbol{\beta}_{i-1} + \boldsymbol{\epsilon}_i, \; i = 2, ...N \tag{4.2.14}$$

where $\mathbf{\Pi_1}$ is the coefficient matrix of dimension $(J_0 \times J_0)$ with eigenvalues of modulus strictly

[4] Aue et al. (2015) show that the one-step ahead forecasts of functional time series from the KL expansion with VAR-based forecasts of eigenscores are asymptotically equivalent to the predictions based on the Functional Autoregressive process of order 1 (FAR(1)) fitted to the functional time series.

less than 1, and $\boldsymbol{\epsilon}_i$ is a multivariate $(J_0 \times 1)$ white noise process with mean zero and variance Σ. Note that the marginal variance of each eigenscore vector being equal to λ_1, implies an additional constraint $\Sigma = \Lambda - \Pi\Pi'$ on the parameters of the model, where Λ is a diagonal matrix with diagonal elements $\lambda_1, ..., \lambda_{J_0}$.

Example 2: KL Conditionally Heteroscedastic Dynamic Factor Model

Let us consider the model of Example 1, with equation (4.2.11) replaced by

$$\beta_{i|J_0} = \Phi\beta_{i-1,J_0} + \epsilon_i, \text{ where } \epsilon_i = H_i^{1/2}\eta_i, \tag{4.2.15}$$

and where H_i is the conditional covariance matrix of $\beta_{i|J_0}$ of dimension $J_0 \times J_0$ given the information set $\mathcal{F}_{i-1}$ generated by the realizations up to and including $i-1$, and η_i is an i.i.d. vector with mean 0 and such that $E\eta\eta' = Id$.

The above dynamic would follow from a decomposition similar to (4.2.13), based on a functional joint FAR-FGARCH model representing jointly the dynamics of conditional mean and variance of return functions, under a relaxed assumption on the errors u_i being $L^4 - m$-approximable (Kokoszka et al., 2017) and with vector $\epsilon_i = (< u_i, \tilde{\xi}_1 >, ..., < u_i, \tilde{\xi}_{J_0} >)'$ of linear transformations of the conditionally heteroscedastic errors. To our knowledge, a joint functional FAR-FGARCH model is not available in the literature. In contrast, the dynamics of the conditional mean and variance can be easily specified for the discrete-time eigenscores in the context of the KL Dynamic Factor model. The empirical counterpart of $\beta_i =< X_i, \hat{\xi}_j >$ can be written as the VAR(1) process in eq. (4.2.14) with weak white noise errors and, for example, the BEKK model of conditional volatility (Engle and Kroner, 1995)[5]. Since the BEKK model gets computationally costly when the number of time series data increases, we can use its variant, called the scalar BEKK (sBEKK(1,1)) (Ding and Engle, 2001):

$$\boldsymbol{H}_i = \boldsymbol{C}\boldsymbol{C}' + a\boldsymbol{\epsilon}_{i-1}\boldsymbol{\epsilon}_{i-1}' + g\boldsymbol{H}_{i-1}, \; i = 1, ...N \tag{4.2.16}$$

[5]Alternatively, other types of models, such as the constant conditional correlation (CCC) model of Bollerslev (1990) and the dynamic conditional correlation (DCC) model of Engle (2002) can be considered.

where, $\boldsymbol{H}_i$ is the $(J_0 \times J_0)$ multivariate conditional variance-covariance matrix of β_i, $a, g \in R$ with $a \geq 0, g \geq 0$ are scalar parameters, and $\boldsymbol{\epsilon}_i$ is the error vector of the VAR(1) representation of eigenscores in equation (4.2.14). Matrix $\boldsymbol{C}$ is a lower triangular matrix of parameters of dimension $J_0 \times J_0$. We assume that the parameters of the sBEKK(1,1) model satisfy the standard stationarity and non-negativity conditions. Since the marginal variance of each eigenscore is equal to λ_i, the parameters of this model are constraint, as illustrated in Appendix B.1.

Example 3: KL Conditionally Heteroscedastic Univariate Dynamic Factor Model

The KL factor model can be simplified further, as in practice one may prefer to model the contemporaneously uncorrelated eigenscore processes using univariate time series models (Shang, 2020; Shang and Kearney, 2022). For conditionally heteroscedastic processes, we can specify the GARCH dynamics for the associated errors as follows.

Let us suppose that the eigenscores $\beta_{i,1}, .., \beta_{i,J_0}$ are independent and follow univariate AR(p)-GARCH(K, L) models. Then, the dynamic of eigenscore $\beta_{i,j}$ can be written as:

$$\begin{aligned} \beta_{i,j} &= \sum_{l=1}^{p} a_{l,j} \beta_{i-l,j} + \epsilon_{i,j}, \\ \nu_{i,j} &= \varsigma_{0,j} + \sum_{k=1}^{K} \varsigma_{k,j} \nu_{i-k,j} + \sum_{l=1}^{L} \zeta_{l,j} \epsilon_{i-l,j}^2 \end{aligned} \tag{4.2.17}$$

where

$$\epsilon_{i,j} = \sqrt{v_{i,j}} z_{i,j}, \quad \forall i = 1, .., N, \tag{4.2.18}$$

and where $a_{l,j}$ is the autoregressive coefficient on lag l of eigenscore j, $\epsilon_{i,j}$ is the error term of eigenscore j, $\varsigma_0, \varsigma_{l,j}$ and $\zeta_{k,j}$ are the coefficients of the conditional variance, and $z_{i,j}$'s are i.i.d. N(0,1) variables (Gourieroux, 2012). We assume that the roots of the autoregressive polynomial $1 - \sum_{l=1}^{p} a_{l,j} x^l = 0$ are outside the unit circle $\forall j = 1, .., J_0$, $\varsigma_0 > 0$, and the remaining coefficients of the GARCH process satisfy the standard stationarity and non-negativity conditions. Note that the marginal variance of each eigenscore being equal to λ_i, implies constraints on the parameters of the model, which are illustrated in Appendix B.1.

To accommodate processes with explosive patterns such as those displayed by the returns on Terra and Luna in the summer of 2022, a heavy-tailed error distribution needs to be assumed, with potentially infinite marginal variance. Then, an IGARCH process could replace the specification given above (Engle and Bollerslev, 1986). These dynamics are out of the scope of this thesis.

The assumption of stationarity of the functional time series in i is crucial for the validity of the proposed method. In practice, it is satisfied by the financial return functions in i (Kokoszka et al., 2014) and can be tested using the approach of Horváth et al. (2014). The assumption of square integrability of each function on $[0, 1]$ is also essential. The FPCA is based on the Karhunen-Loève expansion, which is a spectral decomposition of the covariance operator of the functional data. The existence and properties of this operator depend on the assumption that the random functions are in an L^2 space (space of square-integrable functions), which implies a finite expected variance. If the data-generating process has infinite variance (e.g., heavy-tailed distributions like certain stable laws), the population covariance function does not exist in the traditional sense, and the sample covariance estimates can be unstable and inconsistent, making the principal components unreliable. In practice, the marginal return distributions are characterized by heavy tails. Our approach can accommodate heavy tails but not infinitely heavy tails, using the conditional heteroscedasticity model. Hence, in applied research, it is recommended to verify the finite variance assumption, for example, by estimating the tail index of intraday return distributions.

4.2.2.1 Financial Interpretation

The KL Conditionally Heteroscedastic Dynamic Factor model of return functions can be compared to the ARCH-M model of Engle et al. (1987) for discrete-time observations on returns, formalizing a link between the intraday returns and their conditional daily volatility by introducing a daily risk premium. In our context, the risk premium is $\mu(t)+\sum_{j=1}^{J}\beta_{i,j}\xi_j(t)$, where $\mu(t)$ is the intraday mean curve and the volatility on day i represented by the daily covariance operator, is decomposed into J random daily factors $\beta_{i,j}$ with intraday functional coefficients $\xi_j(t)$. The first factor $\beta_{i,1}$ explains the most intraday conditional volatility, followed by the second factor $\beta_{i,2}$ explaining a smaller portion of conditional volatility, etc. The factor loadings $\xi_j(t)$ accommodate the shape of intraday return functions. Specifically, our eigenfunctions allow us to identify the intraday times of the highest daily risk exposure in Bitcoin, as illustrated later in Section 3. In this respect, our model is similar to the Litterman and Scheinkman (1991) model of bond returns at different maturities that uses principal component analysis in discrete-time data. The KL Dynamic Factor model can also be compared to the Functional Dynamic Factor Model of Kokoszka et al. (2014), for price curves, which is a functional regression model of CIDR of oil futures on the market return and its past value[6].

4.2.2.2 The KL Dynamic Factor-Based Functional Forecast

Let us assume a KL dynamic factor model with mean zero satisfying Definition 1. The theoretical out-of-sample forecast and the conditional variance are derived below in the general case that includes Examples 1,2, and 3.

$$E(X_{N+1}|\underline{X_N}) = E([\sum_{j=1}^{J_0} \beta_{N+1,j}\xi_j]|\underline{X_N}) = \sum_{j=1}^{J_0}[E(\beta_{N+1,j}|\underline{X_N})]\xi_j, \tag{4.2.19}$$

where $\underline{X_N}$ denotes the past curves and $E(\beta_{N+1,j}|\underline{X_N})$ is the forecast of the random eigenscore. By definition, the prediction of $\beta_{N+1,j}$ depends on the past curves through the J_0 first eigenscores only. Therefore, we have:

$$E(X_{N+1}|\underline{X_N}) = \sum_{j=1}^{J_0}[E(\beta_{N+1,j}|\underline{\beta_{N|J_0}})]\xi_j. \tag{4.2.20}$$

It is either based on $E(\beta_{N+1,j}|\underline{\beta_{N,j}})$ when each eigenscore $\beta_{N+1,j},\ j = 1,..,J_0$ is predicted from its own past based on a univariate time series model (Example 2), or $E(\beta_{N+1,j}|\underline{\beta_N})$

[6]As mentioned earlier in the Introduction, our model is also applicable to CIDR, which have a limited economic and financial interpretation (Kokoszka et al., 2014).

where $\beta_N = (\beta_{N,1}, ..., \beta_{N,J_0})$ when the eigenscore j is predicted from its own past and the past values of the remaining $J_0 - 1$ eigenscores (Examples 1 and 3).

The conditional variance at horizon 1 is:

$$V(X_{N+1}|\underline{X_N}) = \sum_{j=1}^{J_0} \sum_{l=1}^{J_0} \xi_j \xi_l' Cov(\beta_{N+1,j}, \beta_{N+1,l}|\underline{\beta_{N|J_0}}) + \omega, \tag{4.2.21}$$

where $\omega = \sigma^2 \sum_{j=J_0+1}^{\infty} \xi_j \xi_j'$, $\sum_{j=J_0+1}^{\infty} \xi_j \xi_j' = Id - \sum_{j=1}^{J_0} \xi_j \xi_j'$ where Id denotes the identity operator on $l_2(I)$ and ξ' denotes the adjoint operator of ξ. Note that $\sum_{j=J_0+1}^{\infty} \xi_j \xi_j'$ and $\sum_{j=1}^{J_0} \xi_j \xi_j'$ are orthogonal projectors (symmetric and idempotent) in the KL dynamic factor model.

4.2.3 New Approach to FPCA Forecasting of Financial Returns

Suppose the returns on Bitcoin (with a fixed holding period) is a stationary time series of functions $X_i(t), t \in I$, each defined over one day and observed on consecutive days $i = 1, ..., N$. Then, the intraday Bitcoin returns can be treated as a continuous time function.

4.2.3.1 Estimation

We approximate the expansions of stochastic functions in (4.2.5) by

$$X_i(t) = \mu(t) + \sum_{j=1}^{J} \beta_{i,j} \xi_j(t) + e_i(t), i = 1, ..., N. \tag{4.2.22}$$

The mean $\mu(t)$ can be consistently estimated from a sample of N functional observations as $\hat{\mu}_N(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t)$. The covariance operator K can be consistently estimated from $\hat{c}_N(x) = \frac{1}{N} < X_i - \hat{\mu}_N, x > (X_i - \hat{\mu}_N)$ (Ramsay and Silverman, 2005b). Aue et al. (2015) show that under general weak dependence assumptions these estimators are $\sqrt{N}$ consistent. Kokoszka et al. (2017) show the consistency and normality of the mean and covariance estimators for functional time series that are autocorrelated and conditionally heteroscedastic.

Under the static FPCA, the functional principal components are extracted from $c(t, s)$, yielding the estimated $\hat{\lambda}_j, \hat{\xi}_j, j = 1, ..., J,$ where $[\hat{\xi}_1(t), \hat{\xi}_2(t), ...]$ are the orthogonal sample eigenfunctions obtained from $\hat{c}(s,t) = \sum_{j=1}^{J} \hat{\lambda}_j \hat{\xi}_j(t) \hat{\xi}_j(s)$ where $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots \geq 0$ are the sample eigenvalues of $\hat{c}(s,t)$ under the static approach. The static FPCA approach has been extended to dynamic FPCA by Aue et al. (2015) for serially correlated functions. Under the dynamic FPCA, the functional principal components are extracted from the long-run covariance matrix defined as $C(t,s) = \sum_{l=-\infty}^{\infty} cov(X_0(s), X_l(t))$, i.e. the marginal covariance of the stationary time series of functions, to ensure the consistency of the estimates. Shang and Kearney (2022) observe that the estimation of dynamic functional principal components using long-run covariance benefits the forecast if there are temporal dependencies in the functional data. Shang (2020) claim that the functional time series method with dynamic functional principal component decomposition does not always outperform that with static functional principal component decomposition.

After the estimation, the in-sample fitted functions $\hat{X}_i(t)$, $i = 1, ..., N$ are obtained from the sample of functional observations $X_1(t)$, ..., $X_N(t)$ as:

$$\hat{X}_i(t) = \hat{\mu}(t) + \sum_{j=1}^{J} \hat{\beta}_{ij} \hat{\xi}_j(t), \; i = 1, ..., N \tag{4.2.23}$$

where $\hat{\mu}(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t)$ is the estimated mean function, $\hat{\beta}_{ij}$ is the j^{th} estimated principal component score for the i^{th} observation where $\hat{\beta}_i =< [X(t) - \hat{\mu}(t)] \hat{\xi}_i(t) >$.

In practice, the optimal choice of truncation J is crucial for the estimation accuracy. The choice criteria are discussed in Shang and Kearney (2022), p.1028 and Shang (2020), p.31. A commonly used approach chooses J set at the minimum, allowing to reach a given fraction δ of the cumulative covariance explained by the first J leading components. For example $J = argmin_{J:J\geq 1} \{ \sum_{j=1}^{J} \hat{\lambda}_j / \sum_{j=1}^{N} \hat{\lambda}_j Id_{\hat{\lambda}_j > 0} \geq \delta \}$, with $\delta = 0.85$.

The variance σ^2 of $e(t)$ can be estimated from the average of estimated daily variances of $\hat{e}_i(t) = X_i(t) - \hat{X}_i(t)$ over N days, where $\hat{X}_i(t)$ is the fitted value defined in eq. (4.2.23). Next, to obtain $\hat{\omega}$, we multiply $\hat{\sigma}^2$ by $diag(Id - \sum_{j=1}^{J} \hat{\xi}_j \hat{\xi}_j')$, based on the entire sample of N

functions to approximate it according to eq. (4.2.21). We expect the result of $Id - \sum_{j=1}^{J} \hat{\xi}_j \hat{\xi}_j'$ to be numerically very close to an identity matrix.

The estimation of the dynamic parameters of eigenscore models depends on the chosen model and is discussed later in this section.

4.2.3.2 Forecast

The forecast at horizon h is computed as follows:

$$\hat{X}_{N+h|N} = \hat{\mu} + \sum_{j=1}^{J} \hat{\beta}_{N+h|N,j} \hat{\xi}_j, \tag{4.2.24}$$

where $\hat{\xi} = [\hat{\xi}_1, ..., \hat{\xi}_J]$ is the set of estimated functional principal components estimated from N functional observations and $\hat{\beta}_{N+h|N,j}$ is the forecast of the j^{th} eigenscore in a univariate representation. The forecast variance is

$$V(\hat{X}_{N+h|N}) = \sum_{j=1}^{J} \sum_{l=1}^{J} \hat{\xi}_j \hat{\xi}_l' \widehat{Cov}(\hat{\beta}_{N+h|N,j}, \hat{\beta}_{N+h|N,l}) + \hat{\omega} \tag{4.2.25}$$

The forecast of the j^{th} eigenscore is obtained from a time series forecasting method as in Aue et al. (2015) and the related papers by Shang (2020) and Shang and Kearney (2022) who documented serial correlation in eigenscore processes $(\beta_{i,j}), i = 1, 2, ...$

4.2.3.3 Forecast Interval Based on Univariate Time Series Model of Eigenscores

In this case, each future eigenscore $\hat{\beta}_{N+1,j}, \ldots, \hat{\beta}_{N+h,j}$ for $j = 1, \ldots, J$ is forecast by using the observations on the past values of each j^{th} eigenscore $\hat{\beta}_{1j}, \ldots, \hat{\beta}_{Nj}$ independently. We estimate the AR(p)-GARCH(K, L) model (4.2.17) in one step by the (quasi) maximum likelihood to forecast each individual eigenscore $\beta_{N+1,j}, \ j = 1, \ldots, J$ and predict each eigenscore along with its future volatility $\hat{\nu}_{N+1|N,j}$. For example, for $h = 1$, the univariate predictions are $\hat{\beta}_{N+1|N,j} = \hat{a}_j \hat{\beta}_{N,j}$ and $\hat{\nu}_{N+1|N,j} = \hat{\varsigma}_{0,j} + \hat{\varsigma}_j \hat{\nu}_{N,j} + \hat{\zeta}_j \hat{\epsilon}_{N,j}^2$ from the AR(1)-GARCH(1,1) model, based on eq. (4.2.17).

The out-of sample forecast of eigenscore $\beta_{N+1,j}$ on day $i = N+1$ and the associated forecast of conditional variance lead to the $(1-\alpha)\%$ pointwise forecast interval calculated as:

$$\hat{X}_{N+1|N}(t) \pm z_{\alpha/2}\sqrt{\sum_{j=1}^{J}\hat{\nu}_{N+1|N,j}\hat{\xi}_j^2(t) + \hat{\omega}(t)} \tag{4.2.26}$$

where $\hat{\nu}_{N+1|N,j}$ is the forecast variance of jth eigenscore estimated by the AR-GARCH for the future function on day $N+1$, $\hat{\xi}_j^2(t)$ is the estimated eigenfunction and $z_{\alpha/2}$ denotes the quantile of the standard Normal distribution.

4.2.3.4 Forecast Interval Based on Multivariate Time Series Model of Eigenscores

For the joint forecast of eigenscores, we apply the sBEKK model to forecast the future conditional volatility of eigenscores ($\beta_{N+1},\ j = 1, \ldots, J_0$) on day $i = N+1$ in two steps. First, we estimate the VAR model (4.2.15) by the (quasi) maximum likelihood and predict the vector of future scores $\hat{\beta}_{N+1|N}$. At horizon $h = 1$ we use the prediction $\hat{\beta}_{N+1|N} = \hat{\Phi}\hat{\beta}_N$ based on equation (4.2.15). In the second step, we apply the sBEKK(1,1) model (4.2.16) to the residuals $\hat{\boldsymbol{e}}_i, i = 1, \ldots, N$ of the VAR(1) model estimated in the first step and forecast the conditional covariance matrix of eigenscores H_{N+1} one day ahead as

$\hat{\boldsymbol{H}}_{N+1|N} = \hat{\boldsymbol{C}}\hat{\boldsymbol{C}}' + a\hat{\boldsymbol{\epsilon}}_N\hat{\boldsymbol{\epsilon}}_N' + g\hat{\boldsymbol{H}}_N.$

The pointwise forecast intervals at level $(1-\alpha)\%$ for the returns one-day-ahead are computed from the diagonal elements of matrix $\hat{\boldsymbol{\xi}}'(t)H_{N+1}\hat{\boldsymbol{\xi}}(t)$ as follows:

$$\hat{\mathbf{X}}_{N+1|N}(t) \pm z_{\alpha/2}\sqrt{diag(\hat{\boldsymbol{\xi}}'(t)\hat{\boldsymbol{H}}_{N+1|N}\hat{\boldsymbol{\xi}}(t)) + \hat{\omega}(t)} \tag{4.2.27}$$

where $\hat{\boldsymbol{X}}_{N+1|N}$ is the forecast one-day-ahead function of returns forecast from the eigenscores $\hat{\beta}_{N+1|N,j},\ \ j = 1, \ldots, J$, estimated by VAR(1)-sBEKK(1,1) model, and $\hat{\boldsymbol{\xi}}(t)$ contains the estimated eigenfunctions $(\hat{\xi}_1(t), \ldots, \hat{\xi}_{J_0}(t))'$

4.2.3.5 Comparison with the Literature

The main difference between our approach compared to Aue et al. (2015), Shang (2020), Shang and Kearney (2022) and Bouri et al. (2021) is that the variance of the forecast jth score is past dependent and predicted, instead of being considered constant and estimated ex-post as in step 4 of the algorithm of Aue et al. (2015). In addition, the parameter estimates of the dynamic model of eigenscores account for conditional heteroscedasticity. Hence, they are more efficient, contributing to more accurate forecasts of eigenscores and then of returns.

In practice, we expect that the FPCA of return functions on Bitcoin generates eigenscores that are serially correlated and conditionally heteroscedastic. Hence, we consider pointwise interval forecasting, which we improve by accommodating the conditional heteroscedasticity of eigenscores. In the existing literature, the forecast interval for FPCA models is obtained from the pointwise forecast bands algorithm. This algorithm, given in Appendix D, was introduced by Aue et al. (2015), and applied by Shang (2020) and Shang and Kearney (2022).

4.3 Data Analysis

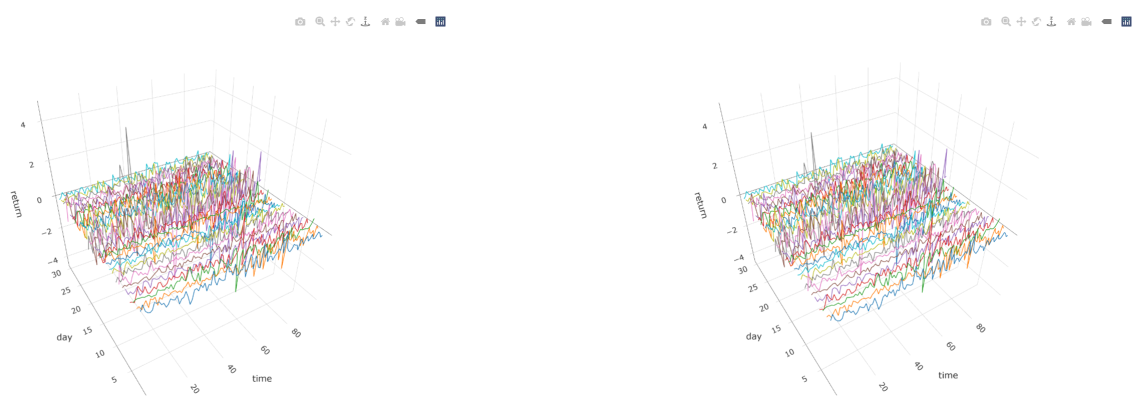
We examine return functions over one UTC day between hours 0:00 and 24:00, and computed from 15-minute and hourly data on prices from Bitsatmp, one of the largest cryptocurrency exchanges, covering the period from January 1, 2022, to December 30, 2023. The 15-minute data comprises 729 days with 96 observations per day, totaling 69984 observations. The hourly data comprise 17472 hourly prices.

Below, we examine the return functions computed from data sampled at 15-minute and hourly, which have financial and economic interpretation. There is a consensus in the literature that, while the time series of price functions are nonstationary, the time series of return functions are stationary, and can be transformed into CIDR through a simple linear transformation. The time series of CIDR functions are considered stationary and examined in e.g. Bouri et al. (2021), Kokoszka et al. (2014).

4.3.1 Intraday 15 Minute Returns

Let us define the 15-minute return on day $i = 1, \ldots, N$ as $X_i^*(t) = (ln(p_{i,t}) - ln(p_{i,t-1})) \times 100$ for $t \in [1, .., 96]$ and $N = 729$. Figure 4.1a illustrates the daily functions of 15-minute BTC returns over days $i = 1, ..., 30$ of the sample.

Figure 4.1: Functions of BTC 15-Minute Returns



(a) 15-Minute Returns

(b) 15-Minute Demeaned Returns

Each line represents one daily function.

We observe that the daily functions of log-returns on consecutive days are not trending, consistent with the stationarity assumption in i. To remove the common intraday mean curve $\mu(t)$, each stochastic function $X_i(t)$, is demeaned as pointed out in Section 2.1, i.e., replaced by $X_i(t) - \mu(t)$, and where $\mu(t) = EX_i(t)$ is independent of index i by the stationarity assumption. Following the approach of Ramsay and Silverman (2005b), each function is demeaned:

$$X_i(t) = X_i^*(t) - \frac{1}{N}\sum_{i=1}^{N} X_i(t),\ t = 1, ..., 96 \tag{4.3.1}$$

by subtracting the daily average 15-minute return at t=1,...,96 over N days. The demeaned return functions, displayed in Figure 4.1b for $i = 1, ..., 30$ show no trend in $i = 1, ...N$.

4.3.2 Intraday Hourly Returns

A similar approach is applied to compute the hourly returns from hourly Bitcoin prices over $N = 729$ days. The returns are defined as $X_i^*(t) = (ln(p_{i,t}) - ln(p_{i,t-1})) \times 100$ for $t \in [1, .., 24]$ and $i = 1, \ldots, N = 729$. Figure 4.1a illustrates the daily functions of 15-minute BTC returns over $i = 1, .., 30$ days.

Figure 4.2: Functions of BTC Hourly Returns

(a) Hourly Returns

(b) Hourly Demeaned Returns

Each line represents one daily function.

As in the previous section, we observe that the log-return functions on consecutive days are not trending, consistent with the stationarity assumption in i. To remove the common intraday mean curve $\mu(t)$, each stochastic function is replaced by $X_i(t) - \mu(t)$, where $\mu(t) = EX_i(t)$. Following the approach of Ramsay and Silverman (2005b), each return function is demeaned:

$$X_i(t) = X_i^*(t) - \frac{1}{N}\sum_{i=1}^{N} X_i(t),\ t = 1, ..., 24 \tag{4.3.2}$$

by subtracting the hourly average returns at $t = 1, ..., 24$ over N days. We observe that the hourly functions of demeaned returns in Figure 4.2b have no trend in $i = 1, ..., N$, similar to 15-minute return functions.

We test the stationarity of the time series of functions using the T_N^0 test of Horváth et al. (2014), implemented e.g. in Wang et al. (2025). The null hypothesis is that the functional time series is stationary in mean and weakly dependent, against the alternatives of a change point in the mean, or a random walk in i. The test results do not reject the null hypothesis at 5% significance level for both 15-minute return functions (p-value = 0.997) and hourly return functions (p-value = 0.658), which suggests that the BTC intraday return functions are indeed stationary.

Moreover, Figures 4.1-4.2 show time-varying volatility of return functions. For example, in 15-minute and hourly return functions, the variation of curves increases on about day 20 and remains high until day 25, suggesting volatility clustering and persistence. We compute the daily realized volatilities from the 15-minute data [see e.g. Andersen and Benzoni (2008)] and display in Figure B.4, Appendix A. We observe that the daily realized volatilities are time-varying and display daily volatility clustering, for example, over days 20-25, 55-65, and 205-215, which is consistent with conditional heteroscedasticity of daily return functions [7].

4.3.3 Forecasts of Daily Return Functions

We consider one-day-ahead forecasts of return functions of 15-minute and hourly returns based on 250 previous daily functions. This approach ensures a sufficient number of in-sample errors allowing us to compare our results with the in-sample error-based forecast intervals of Aue et al. (2015).

We first proceed with the FPCA to compute the eigenscores $\beta_j, j = 1, ..., J$ and eigenfunctions $\xi_j, j = 1, ..., J$. The R-package *fda* is used to perform the FPCA[8]. The estimation of the dynamic parameters of eigenscore models depends on the dynamic model chosen for forecasting of eigenscores, and it is discussed in the next subsection. We determine the optimal number J of eigenfunctions by considering the minimum number of leading components

[7]The results of the commonly used Lagrange Multiplier test reported in Tables B.1 and B.2, Appendix A confirm the presence of conditional heteroscedasticity in discrete-time data at both sampling frequencies.

[8]The Fourier basis functions are selected while performing the FPCA [see Appendix B.2].

that allow us to explain a given level of the cumulative proportion of variance (CPV). We follow Horvath and Kokoszka (2012), who use the criterion

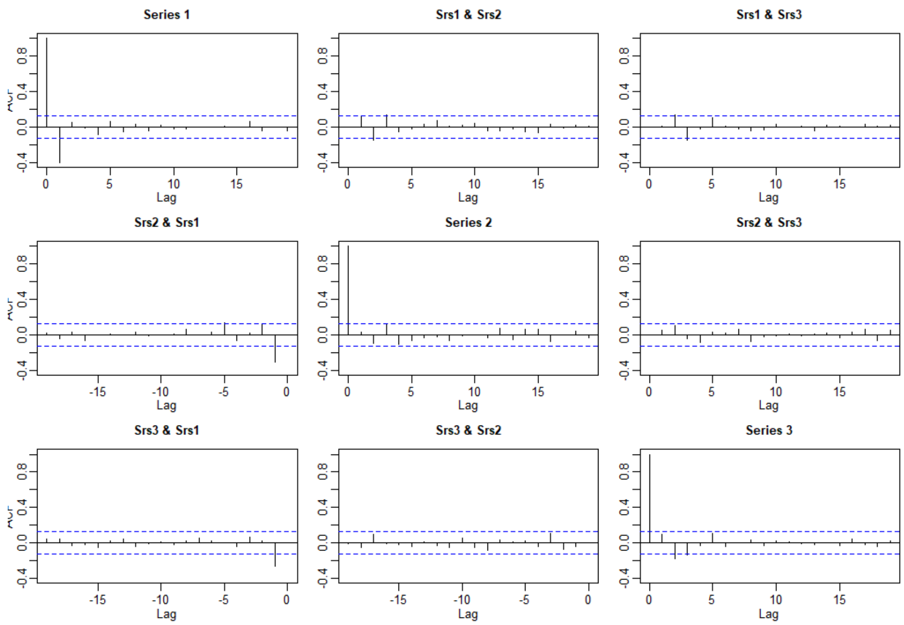
$$J_{\mathrm{CPV}} = argmin_{J:1\leq J\leq L}\left\{\frac{\sum\limits_{j=1}^{J}\hat{\lambda}_j}{\sum\limits_{j=1}^{N}\hat{\lambda}_j} \geq \delta\right\}, \tag{4.3.3}$$

where L is the largest empirical number of eigenfunctions and $\hat{\lambda}_j$ represents the J^{th} estimated eigenvalue and $\delta = 85\%$ (Horvath and Kokoszka, 2012). In our 15-minute data, the first 46 (J=46) eigenvalues explain 85% of variance. In the hourly data, $J = 16$ explain 85% of variance [9].

Next, we compute the auto- and cross-correlation functions of the eigenscores $\beta_j\, j = 1, ..., J$. Figures 4.3 and 4.4 illustrate the autocorrelation functions (ACF) and cross-ACF of the first three eigenscores in 15-minute data.

[9] The choice of 85% follows from experimenting with various thresholds from 50% up to 99%, among which the threshold of 85% is preferred, based on minimizing the interval forecast accuracy score of our proposed models, reported in Tables 4.1 and 4.2 later in this section.

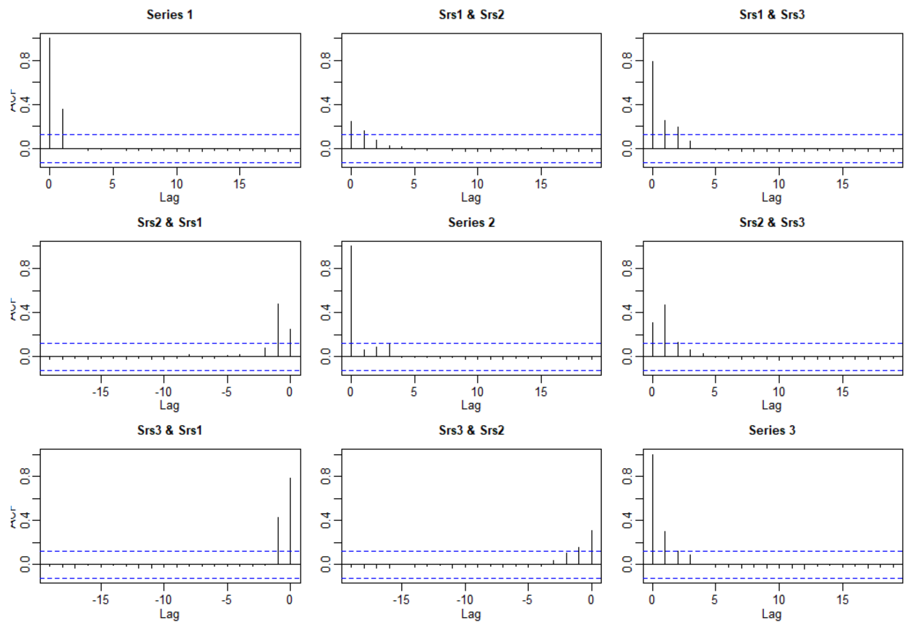
Figure 4.3: ACF of 15-Minute Eigenscores Corresponding with Eigenvalues of Order 1 to 3



The diagonal graphs show the ACF of each eigenscore $\beta_j, j = 1, 2, 3$. The off-diagonal graphs show the cross ACF between the eigenscores β_j, β_k for $j \neq k$.

Figure 4.3 shows that the autocorrelations are statistically significant in all $\beta_j, j = 1, 2, 3$, and in most eigenscores of higher order (not displayed). There also exist statistically significant lagged cross-correlations between β_j, β_k for $j \neq k$, which motivates the use of the VAR model for the conditional means of eigenscores.

Figure 4.4: ACF of 15-Minute Squared Eigenscores Corresponding with Eigenvalues of Order 1 to 3



The diagonal graphs show the ACF of the square of each eigenscore $\beta_j, j = 1, 2, 3$. The off-diagonal graphs show the cross ACF between the squares of eigenscores β_j, β_k for $j \neq k$.

Figure 4.4 shows that the squared eigenscores are serially correlated, and hence the eigenscores are conditionally heteroscedastic. Then, the GARCH effects exist for $\beta_j, j = 1, ...3$ and there are also significant lagged cross-correlations between the squares of β_j, β_k for $j \neq k$ at short lags. The serial correlation in the eigenscores and their squares motivates us to use the AR-GARCH and VAR-sBEKK models of eigenscores.

Figures B.1 and B.2 displaying the ACF for hourly data can be found in Appendix A. Given the risk factor interpretation of eigenscores, we examine the eigenfunctions as the functional factor loadings. The eigenfunctions of 15-minute returns are illustrated in Figures B.5-B.7, Appendix A. We observe that the functional factor loading on the first eigenscore explaining the highest portion of risk on Bitcoin, reveals increased risk exposure at UTC hours 13-14, associated with the opening of the NYSE. The coefficient on the second risk factor points to increased risk exposure about UTC hours 18-19, corresponding to the closure of LSE and other European stock markets. Finally, the third coefficient reveals the effect of NYSE closure at about UTC hour 21 [See, ?]. Hence, these times of a trading day provide higher risk premia and can be exploited by investors to maximize their profits.

4.3.3.1 Performance Measures

The root mean square error (RMSE) and mean absolute error (MAE) are used to evaluate the performance of pointwise forecasts of each future function predicted one-step ahead out of sample. The formulas are:

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{x}_t - x_t)^2} \tag{4.3.4}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{x}_t - x_t| \tag{4.3.5}$$

where n is the number of discrete-time forecast values in a 1-day-ahead function.

The QLIKE(Quasi-Likelihood) loss function is also used in evaluating volatility forecasting (Patton, 2011). It can be written as:

$$L(\hat{\sigma}_t^2, h_t) = \frac{1}{n}\sum_{t=1}^{n}(log\, h_t + \frac{\hat{\sigma}_t^2}{h_t}) \tag{4.3.6}$$

where, the squared return, r_t^2 are used as a volatility proxy of $\hat{\sigma}_t^2$ (Patton, 2011), and h_t denotes the forecast of the conditional variance of forecast returns $\hat{r}_t$. The conditional variance forecasts h_t for Bitcoin returns obtained from the AR(1)-GARCH(1,1) and VAR(1)-sBEKK(1,1) models of eigenscores are computed using eq. (4.2.26) and (4.2.27) respectively. The conditional variance forecast derived from the ARMA(p,q) and VAR(1) model of eigenscores is given by $h_{t,ARMA} = \sum_{j=1}^{J}\hat{\sigma}_{\epsilon_j}^2\xi_{j,t}^2 + \hat{\omega}_t$, where $\hat{\sigma}_{\epsilon_j}^2$ is the conditional variance implied by the fitted residuals of ARMA(p,q) models of eigenscores. For the VAR(1) specification, the conditional variance of Bitcoin returns is $h_{t,VAR} = diag(\hat{\boldsymbol{\xi}}_t \hat{\boldsymbol{\Sigma}}_\epsilon \hat{\boldsymbol{\xi}}_t + \hat{\omega}_t)$, where $\hat{\boldsymbol{\Sigma}}_\epsilon$ is the estimated conditional covariance matrix of the VAR(1) model of eigenscores, obtained from the fitted residuals.

To evaluate the interval forecast accuracy, we use the interval score introduced by Gneiting and Raftery (2007) and Gneiting and Katzfuss (2014). For pointwise forecast intervals at the $100(1-\omega)\%$ nominal coverage probability, the lower and upper bounds at $\omega/2$ and $1-\omega/2$, are denoted by $\hat{X}^{lb}_{N+1}(t)$ and $\hat{X}^{ub}_{N+1}(t)$. The score for the forecast interval is :

$$\begin{aligned} S_\omega[\hat{X}^{lb}_{N+1}(t), \hat{X}^{ub}_{N+1}(t), X_{N+1}(t)] =& [\hat{X}^{ub}_{N+1}(t) - \hat{X}^{lb}_{N+1}(t)] \\ &+ \frac{2}{\omega}[\hat{X}^{lb}_{N+1}(t) - X_{N+1}(t)]\mathbf{1}\{X_{N+1}(t) < \hat{X}^{lb}_{N+1}(t)\} \\ &+ \frac{2}{\omega}[X_{N+1}(t) - \hat{X}^{ub}_{N+1}(t)]\mathbf{1}\{X_{N+1}(t) > \hat{X}^{ub}_{N+1}(t)\} \quad (4.3.7) \end{aligned}$$

where $\mathbf{1}\{\cdot\}$ represents the binary indicator function, and ω is the significance level (Gneiting and Raftery, 2007). The mean score for the one-day-ahead forecast interval is :

$$\bar{S}_\omega = \frac{1}{T}\sum_{t=1}^{T} S_\omega[\hat{X}^{lb}_{N+1}(t), \hat{X}^{ub}_{N+1}(t), X_{N+1}(t)] \quad (4.3.8)$$

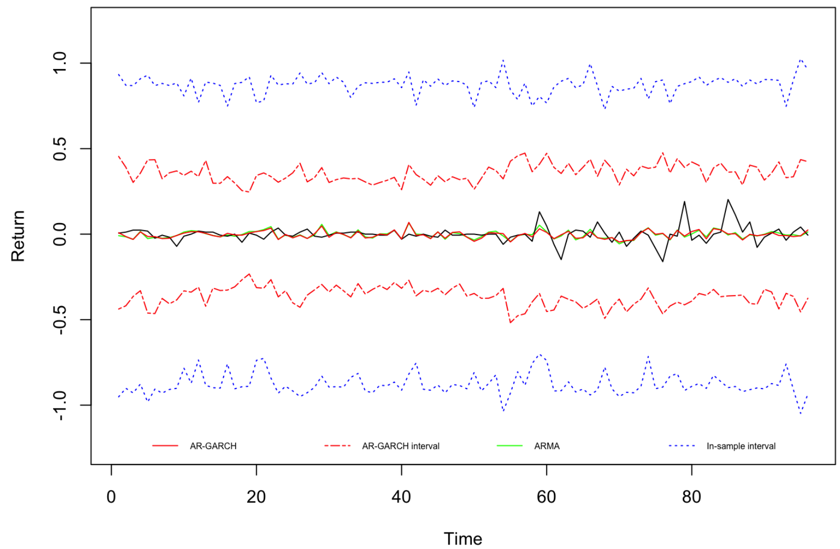
which rewards a narrow forecast interval if, and only if, the true future observation lies within the forecast interval (Shang, 2020).

4.3.3.2 FPCA Forecast with Univariate AR-GARCH Model of Eigenscores

In this section, each eigenscore $\beta_j,\ j = 1, \ldots, J$ is considered an independent time series that follows the AR(1)-GARCH(1, 1) model to account for the conditional heteroskedasticity. Our approach is compared with that Shang and Kearney (2022), where the eigenscores are modeled as ARMA(p,q) processes[10] and the conditional heteroskedasticity is disregarded.

[10] The orders p and q are automatically selected by the auto.arima package in R.

Figure 4.5: One-Day-Ahead 15-Min Return Function Forecast and Forecast Intervals



This figure compares the forecast functions and forecast intervals of 15-Min returns on 2022-12-26 based on AR(1)-GARCH(1,1) and ARMA(p,q) models of eigenscores

We estimate the AR(1)-GARCH(1,1) models (4.2.17) separately for each eigenscore $\beta_{i,j},\ j = 1, \ldots, J$, over days $i = 1, ..., N$, by the quasi maximum likelihood method[11] using the conditional log-likelihood function based on the t-student distribution with estimated degrees of freedom. The autoregressive orders are selected to provide a parsimonious fit to all eigenscores, among which the first three eigenscores display the most serial correlation in levels and squares, as illustrated earlier in this section in Figures 4.3 and 4.4. The degree of freedom estimates accommodate the tails of the conditional densities of $\beta_{i,j}$ for increased flexibility and robustness. The degree of freedom estimates exceed 2.5 in the 15-minute data and 3.2 in the hourly data, indicating that the marginal distributions of eigenscores are heavy-tailed, with finite variance. Then, the eigenscores $\hat{\beta}_{N+1|N,j},\ j = 1, \ldots, J$ are predicted at horizon 1 from the joint AR(1)-GARCH(1,1) model (4.2.17), which also provides the forecasts of conditional volatility of eigenscores $\hat{\nu}_{N+1|N,j}$. The forecast of returns $\hat{X}_{N+1|N}$

[11] The R package *rugarch* is used to estimate the AR-GARCH in this paper.

is obtained from formula (4.2.24) evaluated at the estimated $\hat{\xi}_j$ and the predicted $\hat{\beta}_{N+1|N,j}$. The estimate of $\hat{\omega}$ is based on the remaining 0.15 percent of the total variance of returns, unexplained by $J_{CPV} = 0.85$[12]. Next, the forecasts of $\hat{X}_{N+1|N}$, $\hat{\nu}_{N+1|N,j}$ and $\hat{\xi}$, $\hat{\omega}$ are used to evaluate the expression of the forecast interval in eq. (4.2.26).

We test the null hypothesis of normal distribution of pointwise conditionally studentized forecasts[13] in the 15-minute and hourly functions over 10 days[14]. The Shapiro-Wilk test of the null hypothesis of normal distribution does not reject the null with the p-values of 0.6539 for the 15-minute returns, and 0.9004 for the hourly returns, respectively. Accordingly, the critical value is taken from the normal distribution at $\alpha = 0.05$.

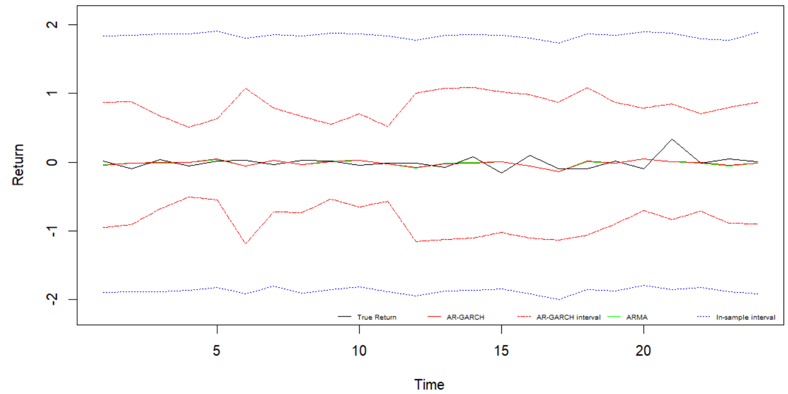
Figures 4.5 and 4.6 show the one-day-ahead forecasts and forecast intervals for the 15-minute and hourly BTC return functions on 2022-12-26, respectively. The one-day-ahead forecast functions, red for AR(1)-GARCH(1,1) and green for ARMA(p,q), are calculated as one-period ahead conditional mean forecasts of AR-GARCH and ARMA(p,q) models, respectively. The AR(1)-GARCH(1,1) forecast intervals of eigenscores are calculated using formula (4.2.26). The ARMA(p,q) forecast intervals are calculated following the algorithm of Aue et al. (2015) given in Appendix D and mentioned in Section 2.3.6.

[12]For illustration, the variance of the forecast error and the volatility of 15-minute returns are plotted in Figure B.4, Appendix A.

[13]We compute pointwise studentized ratios $\hat{X}_{N+1|N,\tau}/\sqrt{\sum_{j=1}^{J}\hat{\nu}_{N+1|N,j}\hat{\xi}^2_{j,\tau}+\hat{\omega}_\tau}$, $\tau = 1, ..., 96$ and $\tau = 1, ..., 24$ for each future daily forecast of 15-minute and hourly data, respectively.

[14]equivalent to 960 and 240 point forecasts.

Figure 4.6: One-Day-Ahead Hourly Return Function Forecast and Forecast Intervals



This figure compares the forecast functions and forecast intervals of hourly returns on 2022-12-26 based on AR(1)-GARCH(1,1) and ARMA(p,q) models of eigenscores

We observe that our forecast function is close to the forecast function of Shang (2020) and Shang and Kearney (2022), and they are both close to the true functions of returns on Bitcoin. We observe a difference in the forecast intervals as those obtained from the AR(1)-GARCH(1,1) model are narrower than those of Aue et al. (2015). This concerns both return series sampled at 15-minute and hourly intervals in Figures 4.5 and 4.6.

Table 4.1: Forecast Performance: AR(1)-GARCH(1,1) and ARMA(p,q) Eigenscore Models

15-Min	**RMSE**	**MAE**	**Sign**	$\bar{S}_\omega$	**Qlike**
AR-GARCH	**0.084**	**0.052**	**47.9 %**	**0.607**	**-3.703**
ARMA	0.086	0.056	46.7%	1.374	-1.710

Hourly	**RMSE**	**MAE**	**Sign**	$\bar{S}_\omega$	**Qlike**
AR-GARCH	**0.168**	**0.115**	**49.7 %**	**1.581**	**-2.041**
ARMA	0.168	0.120	48.2	3.705	-0.856

The results in this table are calculated from 960 point forecasts of 15-minute returns and 240 point forecasts of hourly returns, starting from 2022-12-26.

Table 4.1 shows the performance statistics of one-day-ahead return function forecasts over 10 future days starting from 2022-12-26 computed from the AR(1)-GARCH(1,1) and ARMA(p,q) models of eigenscores. Columns 1 to 6 report the time series model of eigenscores (Col.1), the RMSE (Col.2), MAE (Col.3), the percentage of correctly forecast return signs (Col.4), the interval score (Col.5) and Qlike (Col.6). As pointed out in Patton (2011), MAE may not be reliable, compared to RMSE and QLIKE criteria. The accuracy of pointwise forecasts of return functions from both models is close. We apply the Diebold-Mariano test to the forecast errors of return functions at discrete times $e_t = y_t - \hat{y}_t$, and based on the ARMA(p,q) models and AR(1)-GARCH(1,1) models of eigenscores. Both hourly (p-value of 0.362) and 15-minute (p-value of 0.201) results do not reject the null hypothesis, which suggests that the models have the same forecast accuracy level, which is consistent with the reported RMSE and MAE. Our pointwise forecasts are also similar to the competitor in terms of the correct sign predictions reported in Column 4, as the difference is not statistically significant either.

The difference is in the accuracy of pointwise forecast intervals. The interval scores $\bar{S}_\omega$ (the lower, the better) in Col.5 indicate that the forecast interval estimated by the AR(1)-GARCH(1,1) model improves upon the forecast interval of Aue et al. (2015) based on ARMA(p,q) as in Shang and Kearney (2022) and Shang (2020) for both 15-minute and hourly return functions.

The DM test is applied to the interval forecast scores for both 15-minute and hourly return forecast intervals, which evaluate the accuracy of predictive intervals rather than point forecasts. Both tests applied to 15-minute returns (p-value of 5.754×10^{-5}) and the hourly (p-value of 2.2×10^{-16}) reject the null hypothesis of equal accuracy, implying that the two models exhibit significantly different levels of interval forecast accuracy. This result is consistent with the lower reported QLIKE value of AR-GARCH in Col. 6 of Table 4.1.

Next, we examine empirical coverage of the forecast interval corresponding to a pointwise nominal 95% level ($\omega = 0.05$ in equation (4.3.7)). The empirical coverage rate achieved by our method, indicating the proportion of times true future returns fall within the forecast intervals, is 98.33% for the 15-minute returns and 96.35% for the hourly returns.

4.3.3.3 FPCA Forecast with VAR-sBEKK Model of Eigenscores

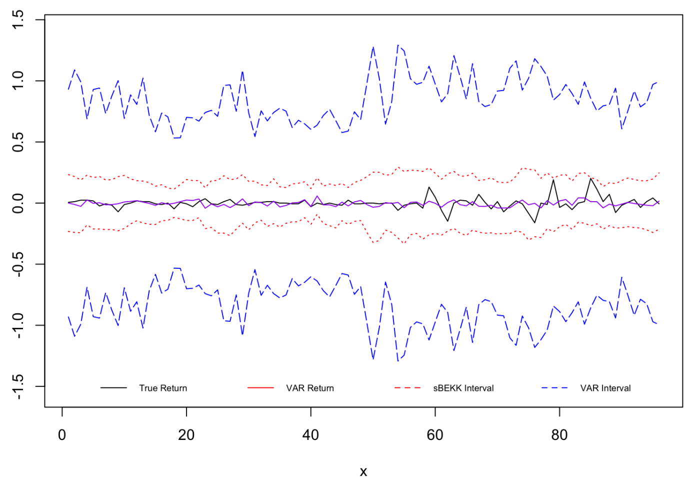
In this section, we take into account the lagged cross-correlations and conditional heteroskedasticity of eigenscores by using the VAR(1)-sBEKK(1,1) models (eq. 4.2.15-4.2.16).

The prediction intervals are computed as follows. The VAR(1)-sBEKK models in equations (4.2.15)-(4.2.16) are estimated from vectors of eigenscores $\beta_i = (\beta_{1,i}, ..., \beta_{J,i})'$ over $, i = 1, ...N$ days by (quasi) maximum likelihood with the conditional log-likelihood function based on the multivariate normal distribution[15], using the approach of Hafner and Herwartz (2008) in two steps. The estimated model (4.2.15)-(4.2.16) provides next the forecasts of eigenscore vector $\hat{\beta}_{N+1|N}$ and conditional volatility matrix $\hat{H}_{N+1|N}$. The predicted eigenscore vector $\hat{\beta}_{N+1|N}$ and matrix $\hat{\boldsymbol{\xi}} = [\hat{\xi}_1 .... \hat{\xi}_J]$ substituted into the multivariate form of equation (4.2.24): $\hat{\boldsymbol{X}}_{N+1|N} = \hat{\mu} + \hat{\boldsymbol{\xi}}\hat{\beta}_{N+1|N}$ provide the forecast of return $\hat{\boldsymbol{X}}_{N+1|N}$. This forecast, along with the predicted matrix of eigenscore volatility $\hat{H}_{N+1|N}$ and matrix $\hat{\boldsymbol{\xi}}$ become inputs in the forecast interval formula (4.2.27). The estimate of $\hat{\omega}$ in the forecast interval is based on the remaining 0.15 percent of the total variance of returns, unexplained by $J_{CPV} = 0.85$. The prediction interval in eq. (4.2.27) is then evaluated using the aforementioned inputs and the critical value of the normal distribution at $\alpha = 0.05$.

Figures 4.7 and 4.8 compare the one-step ahead forecast function and forecast interval of 15-minute and hourly returns based on the VAR(1)-sBEKK(1,1) model (4.2.15)-(4.2.16) of eigenscores with the forecast function and forecast interval obtained from the VAR(1) model following the approach of Aue et al. (2015).

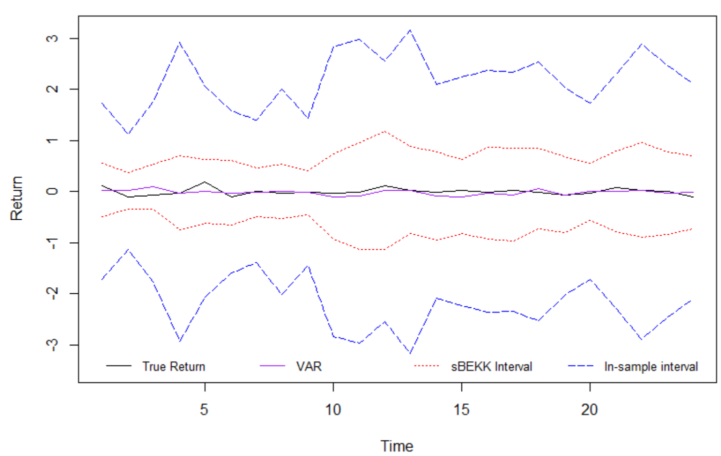
[15] The R package *BEKKs* is used to estimate the sBEKK(1,1) model in this paper.

Figure 4.7: One-Day-Ahead 15-Min Return Function Forecast and Forecast Interval (VAR-sBEKK)



This figure compares the forecast function and forecast interval of 15-Min returns on 2022-12-26 based on the VAR-sBEKK and VAR models of eigenscores

Figure 4.8: One-Day-Ahead 15-Min Return Function Forecast and Forecast Interval (VAR-sBEKK)



This figure compares the forecast function and forecast interval of hourly returns on 2022-12-26 based on the VAR-sBEKK and VAR models of eigenscores

We observe that the proposed method based on the VAR(1)-sBEKK(1,1) model of eigenscores produces forecast intervals that are narrower than those proposed by Aue et al. (2015).

Table 4.2: Forecast Performance: VAR-sBEKK and VAR Eigenscore Models

15-Min	**RMSE**	**MAE**	**Sign**	$\bar{\boldsymbol{S}}_{VB}$	$\bar{\boldsymbol{S}}_{VAR}$	$\bar{\boldsymbol{S}}_{VB}$	$\bar{\boldsymbol{S}}_{VAR}$
	0.089	**0.058**	**47.6 %**	**0.552**	**1.567**	**-4.864**	**-2.932**
Hourly	**RMSE**	**MAE**	**Sign**	$\bar{\boldsymbol{S}}_{VB}$	$\bar{\boldsymbol{S}}_{VAR}$	$Qlike_{VB}$	$Qlike_{VAR}$
	0.185	**0.134**	**45.0%**	**1.443**	**4.115**	**-1.974**	**-1.132**

The results in this table are calculated based on the forecast of the same time interval as in Table 1. $\bar{S}_{VB}$ represents the forecast interval score of the VAR(1)- sBEKK(1,1) model and $\bar{S}_{VAR}$ represents the forecast interval score of Aue et al. (2015).

Table 4.2 describes the performance of pointwise forecasts and interval forecasts over the same time interval as in Section 3.3.2, obtained by applying VAR(1) and VAR(1)-sBEKK(1,1). Columns 1 to 8 report the observed frequency, followed by the RMSE (Col.2), MAE (Col.3), the percentage of correct forecast sign (Col.4), the forecast interval scores of VAR(1)-sBEKK(1,1) and VAR(1) in Col.5 and Col.6, respectively, and the Qlike criterion for VAR(1)-sBEKK(1,1) and VAR(1) in Col.7 and Col.8, respectively. The RMSE and MAE of our method are identical to those based on the VAR(1) model of Aue et al. (2015), because our estimation proceeds in two steps, and the first step produces the conditional mean forecasts. The Sign in Column 4 provides the percentage of successful sign forecast, which are also identical. The advantage of our approach is in the accuracy of the pointwise forecast interval, which can be evaluated by comparing the interval score $\boldsymbol{S_{VB}}$ of our model in Col. 5 to the score $\boldsymbol{S_{VAR}}$ in Column 6 based on the VAR forecast interval of Aue et al. (2015). The results indicate that the VAR-sBEKK model outperforms the VAR-based forecast of Aue et al. (2015) in terms of the accuracy of the forecast intervals in both 15-minute and hourly returns, although the difference is smaller in hourly data. Columns 7 and 8 compare the QLIKE of both methods. The QLIKE of the VAR-BEKK model in Col. 7 is lower than the QLIKE of the VAR model in Col. 8, indicating the advantage of our forecast interval. Again, the difference is smaller in the hourly data.

Our empirical coverage rate, corresponding to the nominal 95% confidence interval ($\omega = 0.05$ in equation (4.3.7)), for the 15-minute returns is 97.91%, whereas for the hourly returns case, it stands at 98.75%.

By comparing the results in Tables 4.1 and 4.2, we find that the multivariate approach based on the VAR-sBEKK model has the best interval accuracy score in 15-minute returns, while the univariate AR-GARCH model has the best interval accuracy score in hourly returns. This finding is consistent with the evidence of more significant lagged cross-correlations of eigenscores in 15-minute data [Figures 4.3-4.4], compared to the hourly returns [Figures B.1-B.2, Appendix A]. In terms of financial interpretation, the proposed forecast intervals do not overestimate the forecast uncertainty.

4.4 Intraday FPCA Forecast

So far, we have considered predicting the entire function of returns out-of-sample one day ahead, i.e., for $i = N + 1$. Let us now consider a new approach for forecasting incomplete functions of returns at shorter intraday horizons of k time units, over 1 or more hours on a given day $i = N$ when the returns are observed only up to a given hour on that day. For assets traded 24/7, the day is conventionally determined by the UTC hours 0:00 and 24:00. A daily function of returns may as well start at a different point and cover 24 consecutive hours. This idea underlies the forecasting approach based on a "rolling FPCA" and is inspired by the paper of Aguilera et al. (1997).

4.4.1 FPCA Rolling Forecast Algorithm

Let us consider the following scenario: On day N, we observe $N-1$ complete functions $X_i(t)$ on the interval $t = [1, \ldots, 24]$, $i = 1, \ldots, N-1$. Suppose that on day N the return function is incomplete: we have observed only the returns up to and including $t = 23$, and we wish to predict the return over the last hour, i.e in the neighborhood of $t = 24$. To do that, we consider an auxiliary set of complete functions $X_i(s)^+$ on the interval $s = [0, \ldots, 23]$, as illustrated in Table 4.3 below, observed on days $i = 1, ..., N$.

First, we perform the FPCA on the set of $N-1$ functions to get the eigenscores $\hat{\beta}_{ij},\ i = 1, \ldots, N-1,\ j = 1, \ldots, J$ and eigenfunctions $\hat{\xi}^*_l(t)$.

Next, we perform the FPCA on the set of N auxiliary functions $X_i(s)^+, s = [0, \ldots, 23],\ i = 1, \ldots, N$, yielding eigenscores $\hat{\alpha}_{ij}, i = 1, \ldots, N,\ j = 1, \ldots, J$, and eigenfunctions $\hat{\xi}^+_j(s)$. In order to predict the neighborhood of $X_N(24)$, we need to predict the eigenscores β_{ij} associated with $X_N(t = 1, \ldots, 24)$. This is done by a regression model.

Table 4.3: Illustration of rolling functions of time

Hour	0(24)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
$X^+(s)$	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	
$X(t)$		•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•

$X^+(s)$ *and* $X(t)$ are both functions on 24 hourly returns with $X^+(s)$ shifted one-hour-behind of $X(t)$

Table 4.3 provides an example of $X^+(s)$ and $X(t)$ when forecasting the return function over hour 24 of UTC time. Note that there are 23 overlapping discrete-time hourly returns between the functions $X^+(s)$ and $X(t)$ [16]. The algorithm proceeds as follows:

Step 1: Calculate N demeaned functions $X_i^+(s)$ by subtracting the sample mean $\bar{X}^+(s) = \frac{1}{N}\sum_{i=1}^{N} X_i^+(s),\ s = 1-k, \ldots, 24-k$.

Step 2: Calculate $N-1$ demeaned functions $X_i(t)$ by subtracting the sample mean $\bar{X}(t) = \frac{1}{N-1}\sum_{i=1}^{N} X_i(t),\ t = 1, ..., 24$.

Step 3: Estimate the eigenfunctions $\hat{\xi}^+_j(s), j = 1, \ldots, J$ from the covariance operator of $X_i^+(s), i = 1, \ldots, N$ and compute eigenscores $\hat{\alpha}_{ij} =< X_i^+, \hat{\xi}^+_j(s) >, i = 1, \ldots, N;\ j =$

[16] For a prediction at horizon $k > 1$, there will be $24 - k$ overlapping discrete-time returns.

$1,\ldots,J$ of $X^+(s)$ by FPCA.

Step 4: Estimate the eigenfunctions $\hat{\xi}_l(t), l = 1,\ldots,J$ from the covariance operator of $X_i(t), i = 1,\ldots,N-1$, and compute the eigenscores $\hat{\beta}_{il} =< X_i, \hat{\xi}_l(t) >$ by FPCA.

Step 5: For each $j = 1,..,J$ run a regression model of a vector of eigenscores $\hat{\beta}_{.j} = [\beta_{1,j},...,\beta_{N-1,J}]'$ of length $N-1$ on the $(N-1)\times J$ matrix $\hat{\alpha} = [\alpha_{1,l},...,\alpha_{N-1,J}]'$, $i = 1,...,N-1$:

$$\hat{\beta}_{.j} = \hat{\alpha}c + v_j, \quad \text{for } j = 1,\ldots,J, \tag{4.4.1}$$

where $c = [c_1,..,c_J]'$ is a vector of coefficients and v_j is the regression error with mean 0 and finite variance.

Step 6: Predict the eigenscores $\beta_{N,j},\ j = 1,\ldots,J$ on day N by using the estimated regression coefficients from **Step 4** and the $1 \times J$ row vector of eigenscores $\hat{\alpha}_N$ estimated in **Step 3** as the explanatory variable:

$$\tilde{\beta}_{N,j} = \hat{\alpha}_N c \ \text{ for } j = 1,\ldots,J, \tag{4.4.2}$$

Step 7: The forecast over the 24th hour on the N^{th} function is obtained from:

$\hat{X}_N(t) = \sum_{j=1}^{J} \tilde{\beta}_{Nj}\hat{\xi}_j(t) + \bar{X}(t)$.

The forecast function $\hat{X}_N(24)$ evaluated at $t = 24$, gives the forecast value of one-hour-ahead target return.

Note that the FPCA performed on $X(t)$ and $X^+(s)$ can differ because the stationarity is satisfied over days $i = 1,..N$. Hence, the covariances and their principal components can vary over the subsets of time t, although this variation is bounded by the square integrability of the functions.

4.4.2 FPCA Rolling Forecast Algorithm: Application to Bitcoin Returns

We forecast the hourly returns on Bitcoin out-of-sample at horizon $k = 1$ of one hour, starting from 2019-02-07, over the same period as subsample 5 in Gradojevic et al. (2023). First, we describe the implementation of the algorithm for forecasting hourly BTC returns. Next, we compare our results with Gradojevic et al. (2023)[17].

4.4.2.1 Algorithm for Hourly Return Forecasts

To determine the best number n of daily functions to initiate the algorithm, we apply steps 1 to 7 of the algorithm to predict returns in sample using the fixed windows of $n = 90, 91, \dots,$ up to $n = 120$ past functions. The results indicate that using different numbers n in the window only changes the MSE slightly. However, the number of past functions n does affect the correct forecast sign rate. Therefore, the number of past functions n used in our algorithm is chosen to maximize the percentage of correct in-sample forecast signs.

In Step 5, we expect the off-diagonal elements of the $\alpha'\alpha$ matrix to be close to zero by the orthogonality conditions, and the diagonal elements for large j to be close to zero as well. Therefore, we use regularized OLS estimators of coefficients c, such as the Lasso and Ridge estimators[18]. The Support Vector Machine (SVM) is considered too, as it provides optimal weighting and bias correction in addition to accounting for nonlinearities in eq. (4.4.1-4.4.2). We also account for potential non-linearities of these eigenscore relations by using the Random Forest (RF), and Neural Networks (NN) with α's as inputs and β' as outputs.

We briefly describe the setup of the above machine learning methods used in Step 5, refering to β's as outputs and α's as inputs.

The Support Vector Machine (SVM) method is a supervised learning method designed

[17] In this section, the BTC returns are defined as $ln(p_t) - ln(p_{t-1})$, without factor 100 to facilitate this comparison.

[18] The ten-fold cross-validation is used to determine the penalty term in the Lasso and Ridge regressions.

to handle nonlinear relationships between predictors and the target variable (James et al., 2021). It can also be referred to as Support Vector Regression (SVR) when dealing with regression tasks. The R package *e1071* is used in this study to implement SVR (Meyer et al., 2024). The entire series of eigenscores $\hat{\alpha}_{ij}, i = 1, ..., N-1$ estimated in **Step 3** and corresponding series of eigenscores $\hat{\beta}_{il},\ i = 1, ..., N-1$ estimated in **Step 4** are used as the training set. Parameters *cost* and *gamma* are chosen by a 10-fold cross-validation.

The Random Forest (RF) is implemented from the R package *randomForest* (Liaw and Wiener, 2002) and applied to the same inputs and outputs as in SVM. The random forest is built using 100 trees, with the number of candidate variables considered at each split set to the largest integer less than or equal to $log_2 J + 1$, following Gradojevic et al. (2023).

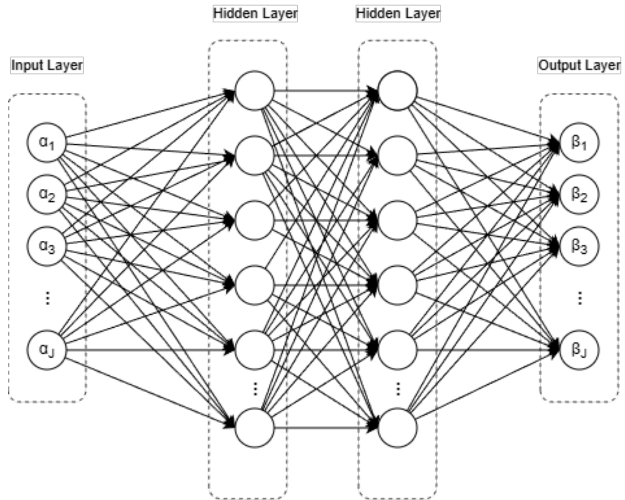
The SVM and RF estimators focus on estimating a single output variable using multiple regressors. In contrast, the feed-forward neural networks (FNN) estimator employs a neural network that simultaneously incorporates multiple input variables and produces multiple output variables.

For notational simplicity, the structure of a single-hidden-layer ANN can be written as:

$$\beta_j = f\Big(b_0 + \sum_{k=1}^{K} b_k Z_k\Big) \text{ where, } Z = g(w_0 + \sum_{j=1}^{J} w_j \alpha_j) \tag{4.4.3}$$

where b_k and w_j are the weights connecting adjacent layers, and K is the number of nodes inside the hidden layer. The terms b_0 and w_0 correspond to the bias parameters, while $f(\cdot)$ and $g(\cdot)$ denote predetermined activation functions. With multiple hidden layers (a multilayer perceptron, MLP), additional nonlinear transformations are composed sequentially, enabling the network to approximate complex nonlinear relationships.. Figure 4.9 illustrates an MLP with two hidden layers used in our study.

Figure 4.9: Four Layer Artificial Neural Network



This figure shows an example of a multilayer perceptron (MLP) with two hidden layers applied to the proposed structure of step 5.

Since neural networks typically perform poorly with small sample sizes, we increase the number of past functions to $n = 800$ to provide sufficient training data. We implement the neural network models in R using the *keras* and *tensorflow* packages (Allaire et al., 2024b,a). The sigmoid activation functions are implemented in the hidden layers, and a linear activation function is used in the output layer. The number of nodes inside the hidden layers is set as $K = 64$. To mitigate overfitting, the following commonly used forms of regularization are applied Gradojevic et al. (2023):

- 10% of the training sample is held out as validation data;
- early stopping halts training if the validation loss (based on RSS) fails to improve for 10 epochs;
- a dropout rate of 0.1 randomly removes hidden units during training.

The results obtained from applying different estimators/predictors in step 5 to equations (4.4.1-4.4.2), which are the OLS, regularized OLS, and nonlinear approximations of equations (4.4.1-4.4.2) based on SVM, RF and NN are reported in Table 4.4 below.

Table 4.4: Forecast Performance with Different Estimators/Predictors in Step 5

Estimator	RMSE $\times 10^{-2}$	Sign (%)
OLS	0.380	59.5
Ridge OLS	0.370	60.5
LASSO OLS	0.370	62
SVM	0.320	57
RF	0.338	58.5
NN	0.331	53.5

The results in this table are calculated from 200 point forecasts of hourly returns starting from 2019-02-27.

Our proposed algorithm gives the best performance in terms of forecast error variance when the SVM method is used in Step 5, which yields an RMSE of 0.0032 while the LASSO OLS results in the best forecast sign rate of 62.5%. To determine if the differences between the methods used in step 5 are statistically significant, the Diebold-Mariano test is applied to each pair of forecasts. We consider the null hypothesis: "the two estimators have equal predictive accuracy", and the alternative hypothesis: "the two estimators have unequal predictive accuracy" and report the p-values in Table 4.5.

Table 4.5: DM Test: P-values for Pairs of Different Estimators

	OLS	Ridge OLS	Lasso OLS	SVM	RF	NN
OLS	NA	0.009	0.003	0.098	0.441	0.059
Ridge OLS	0.009	NA	0.174	0.002	0.029	0.002
Lasso OLS	0.003	0.174	NA	< 0.001	0.022	< 0.001
SVM	0.098	0.002	< 0.001	NA	0.017	0.849
RF	0.441	0.029	0.022	0.017	NA	0.022
NN	0.059	0.002	< 0.001	0.849	0.022	NA

We conclude that the predictive accuracy of the SVM, RF and NN method used in step 5 is significantly different from that of the Ridge and Lasso OLS. The difference between SVM and NN is not statistically significant.

McNemar's test evaluates whether two forecasting models differ significantly in their directional accuracy by testing whether the counts of discordant forecast outcomes (A correct & B incorrect vs. A incorrect & B correct) are symmetric. The null hypothesis suggests no significant difference between the two forecasting models' directional accuracy. A rejection of the null indicates that the two models do not achieve the same rate of correct directional predictions (Diebold, 2015).

Table 4.6: McNemar's Test: P-values for Pairs of Different Estimators

	OLS	Ridge OLS	Lasso OLS	SVM	RF	NN
OLS	NA	0.004	0.004	0.421	0.003	1
Ridge OLS	0.004	NA	1	0.036	0.912	0.002
Lasso OLS	0.004	1	NA	0.056	0.817	0.002
SVM	0.421	0.036	0.056	NA	0.019	0.348
RF	0.003	0.912	0.817	0.019	NA	0.001
NN	1	0.002	0.002	0.348	0.001	NA

Table 4.6 reports the p-values of the McNemar test applied to the forecast sign accuracy. We conclude that the LASSO OLS sign accuracy is statistically different from all methods except for Random Forest (RF), and marginally SVM at 5% significance level.

4.4.2.2 Comparison with Gradojevic et al. (2023)

We compare our out-of-sample forecast performance with Gradojevic et al. (2023), who use nonlinear predictive models of Feedforward Deep Artificial Neural Network(FF-D-ANN), SVM, and RF to forecast discrete-time observations on hourly returns from more than $K = 20$ exogenous variables, which are the technical indicators such as market trend, momentum, and other variables, such as sentiment and volume. These variables are used as inputs in the above machine learning methods. Next, the forecasts produced by these nonlinear models are compared to the linear Random Walk (RW) and ARMAX(1,1). The random walk in the context of returns is a regression of BTC returns on a constant. The ARMAX is an Autoregressive Moving Average with the exogenous variables used as additional regressors. As mentioned earlier, this study of one-hour-ahead BTC returns concerns the same time period (subsample 5) as ours.

In the FF-D-ANN estimation, Gradojevic et al. (2023) consider model (36) with Bitcoin returns as outputs and K exogenous variables as inputs. They employ 3 hidden layers, with 100, 30 and 10 neurons. The training algorithm is Adam, with a sigmoid activation function, and the training period is equivalent to 90% of the data. The training period equivalent to 90% of the data is also used in the application of the SVM and RF methods. The SVM method of prediction and classification used in Gradojevic et al. (2023) relies on transforming the training data into a space of higher dimension using a non-linear mapping, with a polynomial kernel function defined as the dot product of K exogenous input variables. In the second step, the loss function is minimized between linear regressions of training data and targets using a deviation parameter not greater than 0.001. The Random Forest parameters are selected by Gradojevic et al. (2023) by keeping the subsets of inputs of the size of the first integer lower than $log2K + 1$, where K is the number of inputs, while the number of trees adopted to build the RF structure is 100.

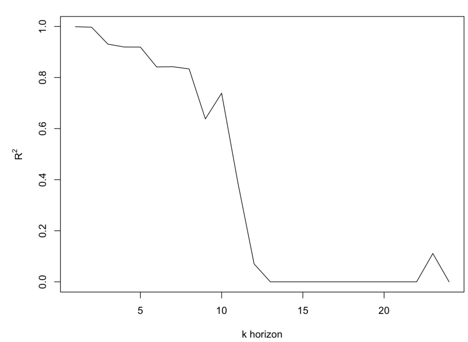
Their conclusion is the relative inability of all nonlinear models to statistically significantly improve upon the random walk model. The random walk model is therefore considered the best predictor of Bitcoin returns. In Gradojevic et al. (2023), the Random Walk model performs best, resulting in an RMSE of 0.00363 and a correct forecast sign rate of 50.24%.

We observe that the predictive accuracy of our method with the SVM, RF and NN methods used in step 5 is statistically significantly different from the predictive accuracy of the best performing random walk model reported by Gradojevic et al. (2023) in terms of the RMSE. In addition, the LASSO OLS provides a statistically significantly better return sign forecast. Overall, in terms of accuracy and sign combined, the best result is obtained from RF, which performs statistically significantly better than the competitor, in terms of forecast and sign accuracy.

4.4.3 FPCA Rolling Forecast Algorithm: the Horizon Effect

So far, we have discussed the forecasts at horizon $k = 1$. At $k = 1$, the coefficients α are close to β, so that the coefficient of determination R^2 of the Ridge, or Lasso regression in step 5 is close to 1. When the forecast horizon k increases, the auxiliary set of functions $X^{+}(s)$ gets shifted backward in time by k and the overlap between $X(t)$ and $X^{+}(s)$ diminishes. Then, the explanatory power of coefficients α decreases with the horizon k and the R^2 of the regression in step 5 diminishes quickly. We observe that the R^2 is above 0.5 up to about horizon $k = 10$, as illustrated in Figure 4.10 below.

Figure 4.10: R^2 for Different Horizons of Using LASSO In Step 5



This figure shows the R^2 of LASSO estimator as the horizon k increases.

Hence, we recommend using the above intraday forecasting method at short horizons, especially at $k = 1$.

4.5 Conclusions

Accurate financial return forecasts are crucial for investors to make informed decisions, manage risk, and plan for their financial goals. Based on accurate return forecasts, investors can better determine if a potential investment aligns with their financial objectives, adjust their strategies accordingly, and prepare for worst-case scenarios by creating contingency plans to mitigate the risk. It is particularly important in the context of trading a highly volatile asset, such as Bitcoin.

In this paper, we define the KL Conditional Heteroscedastic Dynamic Factor model for the analysis of conditionally heteroscedastic functional processes. It accommodates the time-varying volatility of returns and intra-day risk exposure. Based on that model, we introduce two new FPCA methods for forecasting intraday Bitcoin return. The first approach aims at improved interval forecasting of daily return functions and exploits the serial correlation of eigenscores revealed in the FPCA literature. The novelty of our approach is that it takes into consideration the conditional heteroscedasticity of returns to produce improved pointwise forecast intervals. The second approach is a rolling FPCA algorithm for intraday forecasting of incomplete functions of Bitcoin returns at selected points of time. It considers a sequence of partially overlapping daily return functions that start at subsequent time points and allows us to forecast the returns intra-daily.

The empirical results of forecasting Bitcoin daily 15-minute and hourly return functions from the proposed methods indicate that accounting for the conditional heteroscedasticity of returns by forecasting the eigenscores using AR(1)-GARCH(1,1) and VAR(1)-sBEKK(1,1) models noticeably improves the accuracy of forecast intervals. Our proposed "rolling" FPCA method for forecasting one-step-ahead hourly Bitcoin returns shows good performance in terms of both forecast errors and forecast sign accuracy compared to other methods of Bitcoin forecasting considered by Gradojevic et al. (2023) in the recent literature.

Chapter 5

Conclusion

This thesis investigates the dynamics of cryptocurrency markets from both empirical and methodological perspectives, with particular focus on high-frequency intraday and intraweek behavior. The continuous trading of cryptocurrency and common periodic patterns motivate the use of functional data analysis for intraday return processes. By treating high-frequency observations as realizations of underlying functions, the thesis uncovers systematic market structure and assesses exposure to systemic risk of cryptocurrency. Furthermore, this thesis also develops tools for forecasting complete and partial future return functions.

Chapters 2 and 3 document systematic evidence of common intraday and intraweek patterns in prices, returns, volumes, and volatility across a wide range of cryptocurrencies. By modeling high-frequency returns as functional observations rather than discrete data, this study uncovers meaningful temporal structures that are not captured by standard discrete time series methods. The results show that the exposure of native cryptocurrencies and tokens to systemic risk displays intraday and intraweek patterns. On the other hand, stablecoins are immune to systemic risk with distinct dynamics consistent with their institutional design.

Chapter 4 introduces the Karhunen-Loeve dynamics factor model as a methodological contribution of this thesis for the analysis of conditionally heteroscedastic functional processes of intraday Bitcoin returns. Building on this framework, that chapter develops two novel FPCA-based forecasting approaches. The first approach delivers improved pointwise interval forecasts for intraday return functions, while the second proposes a rolling FPCA algorithm that enables forecasting in the presence of incomplete functional observations. Together, these developments highlight the practical and theoretical potential of functional factor models for high-frequency cryptocurrency analysis.

By documenting common intraday and intraweek dynamics across cryptocurrencies and developing functional factor models that incorporate conditional heteroskedasticity, this thesis reveals that both return dynamics and uncertainty evolve in a structured, time-varying manner. These results highlight the advantages of modeling intraday data in the functional domain and underscore the relevance of functional factor models to understand risk, serial dependence, and predictability in digital asset markets.

Bibliography

Aguilera, A. M., Ocaña, F. A., and Valderrama, M. J. (1997). "An Approximated Principal Component Prediction Model for Continuous-Time Stochastic Processes". *Applied Stochastic Models and Data Analysis*, 13(2):61–72.

Ahmed, W. M. and Al Mafrachi, M. (2021). "Do Higher-order Realized Moments Matter for Cryptocurrency Returns?". *International Review of Economics & Finance*, 72:483–499.

Allaire, J., RStudio, and Google (2024a). *"tensorflow: R Interface to 'TensorFlow'"*. R package version 2.16.0.

Allaire, J. J., Tang, Y., and Chollet, F. (2024b). *"keras: R Interface to 'Keras'"*. R package version 2.12.0.

Andersen, T. G. and Benzoni, L. (2008). "Realized Volatility". *IDEAS Working Paper Series from RePEc.*

Aue, A., Horváth, L., and F. Pellatt, D. (2017). "Functional Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Time Series Analysis*, 38(1):3–21.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). "On the Prediction of Stationary Functional Time Series". *Journal of the American Statistical Association*, 110(509):378–392.

Avellaneda, M., Healy, B., Papanicolaou, A., and Papanicolaou, G. (2022). "Principal Eigenportfolios for U.S. Equities". *SIAM Journal on Financial Mathematics*, 13(3):702–744.

Benatia, D., Carrasco, M., and Florens, J.-P. (2017). "Functional Linear Regression with Functional Response". *Journal of Econometrics*, 201(2):269–291.

Benko, M., Härdle, W., and Kneip, A. (2009). "Common Functional Principal Components". *The Annals of Statistics*, 37(1):1 – 34.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). "Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors". *Journal of Financial Economics*, 104(3):535–559.

Bitstamp (2022). Bitstamp cryptocurrency exchange. https://www.bitstamp.net. Accessed 2026-06-09.

Bollerslev, T. (1990). "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model". *The Review of Economics and Statistics*, 72(3):498–505.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer New York, NY.

Bouri, E., Azzi, G., and Dyhrberg, A. H. (2017). "On the Return-Volatility Relationship in the Bitcoin Market Around the Price Crash of 2013". *Economics*, 11(1):2.

Bouri, E., Lau, C. K. M., Saeed, T., Wang, S., and Zhao, Y. (2021). "On the Intraday Return Curves of Bitcoin: Predictability and Trading Opportunities". *International Review of Financial Analysis*, 76:101784–.

Caporale, G. M. and Plastun, A. (2019). "The Day of the Week Effect in the Cryptocurrency Market". *Finance Research Letters*, 31.

Diebold, F. X. (2015). *Forecasting: Principles and Practice*. Edward Elgar Publishing, Cheltenham, UK.

Ding, Z. and Engle, R. F. (2001). "Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing". *Academia economic papers*, 29(2):157–184.

Dunbar, K. and Owusu-Amoako, J. (2022). "Cryptocurrency Returns under Empirical Asset Pricing". *International Review of Financial Analysis*, 82:102216.

Engle, R. (2002). "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models". *Journal of business & economic statistics*, 20(3):339–350.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50.

Engle, R. F. and Kroner, K. F. (1995). "Multivariate Simultaneous Generalized ARCH". *Econometric theory*, 11(1):122–150.

Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model". *Econometrica*, 55(2):391–407.

Eross, A., McGroarty, F., Urquhart, A., and Wolfe, S. (2019). "The Intraday Dynamics of Bitcoin". *Research in International Business and Finance*, 49:71–81.

Fraz, A., Hassan, A., and Chughtai, S. (2019). "Seasonality in Bitcoin Market". *NICE Research Journal*, pages 1–11.

Gneiting, T. and Katzfuss, M. (2014). "Probabilistic Forecasting". *Annual Review of Statistics and Its Application*, 1(1):125–151.

Gneiting, T. and Raftery, A. E. (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". *Journal of the American Statistical Association*, 102(477):359–378.

Gourieroux, C. (2012). *ARCH Models and Financial Applications*. Springer Series in Statistics. Springer, New York.

Gourieroux, C. and Jasiak, J. (2001). *Financial Econometrics; Problems, Models and Methods*. Princeton University Press.

Gradojevic, N., Kukolj, D., Adcock, R., and Djakovic, V. (2023). "Forecasting Bitcoin with Technical Analysis: A Not-So-Random Forest?". *International Journal of Forecasting*, 39(1):1–17.

Grobys, K., Junttila, J., Kolari, J. W., and Sapkota, N. (2021). "On the Stability of Stablecoins". *Journal of Empirical Finance*, 64:207–223.

Gyamerah, S. A. and Cai, N. (2021). "Two-Stage Hybrid Machine Learning Model for High-Frequency Intraday Bitcoin Price Prediction Based on Technical Indicators, Variational Mode Decomposition, and Support Vector Regression". *Complexity (New York, N.Y.)*, 2021(1).

Hafner, C. M. and Herwartz, H. (2008). "Analytical Quasi Maximum Likelihood Inference in Multivariate Volatility Models". *Metrika*, 67(2):219–239.

Hansen, P. R., Kim, C., and Kimbrough, W. (2024). "Periodicity in Cryptocurrency Volatility and Liquidity". *Journal of Financial Econometrics*, 22(1):224–251.

Hobbs, B. F., Helman, U., Jitprapaikulsarn, S., Konda, S., and Maratukulam, D. (1998). "Artificial Neural Networks for Short-Term Energy Forecasting: Accuracy and Economic Value". *Neurocomputing (Amsterdam)*, 23(1):71–84.

Horváth, L., Kokoszka, P., and Rice, G. (2014). "Testing Stationarity of Functional Time Series". *Journal of Econometrics*, 179(1):66–82.

Horvath, L. and Kokoszka, P. P. (2012). *Inference for Functional Data with Applications*. Springer series in statistics. Springer, New York, 1st ed. 2012. edition.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, 2nd edition.

Jasiak, J. and Zhong, C. (2024). "Intraday and Daily Dynamics of Cryptocurrency". *International Review of Economics and Finance*, 96:103658.

Katsiampa, P. (2017). "Volatility Estimation for Bitcoin: A Comparison of GARCH models". *Economics Letters*, 158:3–6.

Kinateder, H. and Papavassiliou, V. G. (2021). "Calendar Effects in Bitcoin Returns and Volatility". *Finance Research Letters*, 38:101420.

Kokoszka, P., Miao, H., and Zhang, X. (2014). "Functional Dynamic Factor Model for Intraday Price Curves". *Journal of Financial Econometrics*.

Kokoszka, P., Rice, G., and Shang, H. L. (2017). "Inference for the Autocovariance of a Functional Time Series Under Conditional Heteroscedasticity". *Journal of Multivariate Analysis*, 162:32–50.

Liaw, A. and Wiener, M. (2002). *"randomForest: Breiman and Cutler's Random Forests for Classification and Regression"*. R package version 4.7-1.1.

Litterman, R. and Scheinkman, J. (1991). "Common Factors Affecting Bond Returns". *The Journal of Fixed Income*, (1):456–477.

Liu, C., Ray, S., and Hooker, G. (2017). "Functional Principal Component Analysis of Spatially Correlated Data". *Statistics and Computing*, 27:1639–1654.

Loeve, M. (1978). *"Probability Theory, Vol 2", Graduate Texts in Mathematics, Vol 46 (4th ed.)*. Springer Verlag.

Long, H., Zaremba, A., Demir, E., Szczygielski, J. J., and Vasenin, M. (2020). "Seasonality in the Cross-Section of Cryptocurrency Returns". *Finance Research Letters*, 35:101566.

Ma, D. and Tanizaki, H. (2019). "The Day-Of-The-Week Effect on Bitcoin Return and Volatility". *Research in International Business and Finance*, 49:127–136.

Makridakis, S., Hogarth, R. M., and Gaba, A. (2009). Forecasting and Uncertainty in the Economic and Business World. *International Journal of Forecasting*, 25(4):794–812.

Mehta, J. and Gurland, J. (1969). "Testing Equality of Means in the Presence of Correlation". *Biometrika*, 56(1):119–126.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2024). *"e1071: Misc Functions of the Department of Statistics, TU Wien"*. R package version 1.7-14.

Patton, A. J. (2011). "Volatility Forecast Comparison Using Imperfect Volatility Proxies". *Journal of Econometrics*, 160(1):246–256.

Prono, T. (2015). "Market Proxies as Factors in Linear Asset Pricing Models: Still Living with The Roll Critique". *Journal of Empirical Finance*, 31:36–53.

Ramsay, J. and Silverman, B. (2005a). *"Functional Data Analysis"*. Springer New York, NY, 2nd edition.

Ramsay, J. and Silverman, B. (2005b). *Functional Data Analysis*. Springer.

Rodrigues, F. and Machado, M. (2025). "High-Frequency Cryptocurrency Price Forecasting Using Machine Learning Models: A Comparative Study". *Information (Basel)*, 16(4):300–.

Sehra, A., Cohen, R., and Arulchandran, V. (2018). "On Cryptocurrencies, Digital Assets and Private Money". *Journal of Payments Strategy & Systems*, 12(1):13–32.

Shahzad, S. J. H., Bouri, E., Ahmad, T., Naeem, M. A., and Vo, X. V. (2021). "The Pricing of Bad Contagion in Cryptocurrencies: A Four-Factor Pricing Model". *Finance Research Letters*, 41:101797.

Shang, H. L. (2020). "Dynamic Principal Component Regression for Forecasting Functional Time Series in A Group Structure". *Scandinavian Actuarial Journal*, 2020(4):307–322.

Shang, H. L. and Kearney, F. (2022). "Dynamic Functional Time-series Forecasts of Foreign Exchange Implied Volatility Surfaces". *International Journal of Forecasting*, 38(3):1025–1049.

Shen, D., Urquhart, A., and Wang, P. (2020). "A Three-Factor Pricing Model for Cryptocurrencies". *Finance Research Letters*, 34:101248.

Simmet, A. and Pohlmeier, W. (2020). "The CAPM with Measurement Error: ' There's life in the old dog yet!'". *Jahrbücher für Nationalökonomie und Statistik*, 240(4):417–453.

Tavares, R. d. S., Caldeira, J. F., and Raimundo Júnior, G. d. S. (2021). "Cryptocurrencies: Formation of Returns from the CRIX Index". *Applied Economics Letters*, 28(8):691–695.

TradingView (2026). Total: Global market cap index (60m). https://www.tradingview.com/symbols/TOTAL/?timeframe=60M.

Vidal-Tomás, D. (2022). "Which Cryptocurrency Data Sources Should Scholars Use?". *International Review of Financial Analysis*, 81:102061–.

Wang, D., Lu, Z., Liu, Z., Xue, S., Guo, M., and Hou, Y. (2025). "A Functional Mixture Prediction Model for Dynamically Forecasting Cumulative Intraday Returns of Crude Oil Futures". *International Journal of Forecasting*.

Wen, Z., Bouri, E., Xu, Y., and Zhao, Y. (2022). "Intraday Return Predictability in the Cryptocurrency Markets: Momentum, Reversal, or Both". *The North American Journal of Economics and Finance*, 62:101733–.

Yarovaya, L. and Zięba, D. (2022). "Intraday Volume-Return Nexus in Cryptocurrency Markets: Novel Evidence From Cryptocurrency Classification". *Research in International Business and Finance*, 60:101592–.

119

Appendix A

Technical Appendix for Chapters 2 & 3

Table A.1: Hourly Data Summary of Cryptocurrencies Traded on Bitstamp

Type	Name	Symbol	No. of zero volumes	No. of Non-zero volumes	Average volume	% Non-zero volumes
Native Coins						
	Bitcoin	BTC	1	4295	101.34	99.98
	Stellar	XLM	1	4295	339781.2	99.98
	XRP	XRP	1	4295	1532195	99.98
	Ethereum	ETH	2	4294	703.12	99.95
	Litecoin	LTC	2	4294	934.12	99.95
	Bitcoin Cash	BCH	3	4293	242.5	99.93
	Cardano	ADA	471	3825	17301.84	89.04
	Algorand	ALGO	534	3762	12247.75	87.57
	Songbird	SGB	965	3331	50529.02	77.54
	Hedera Hashgraph	HBAR	986	3310	14643.05	77.05
Stable Coins						
	Tether	USDT	342	3954	87010.29	92.04
	USD Coin	USDC	581	3715	75208.69	86.48
	Dai	DAI	3332	964	1518.4	22.44
	Paxos Standard Token	PAX	3700	596	529.31	13.87
	Gemini Dollar	GUSD	4178	118	71.01	2.75
	Tether EURt	EURT	4218	78	26.59	1.82
Token						
	Chainlink	LINK	5	4291	6397.82	99.88
	Ploygon	MATIC	320	3976	17771.6	92.55
	Uniswap	UNI	461	3835	1324.7	89.27
	AAVE	AAVE	588	3708	66.77	86.31
	The Sandbox	SAND	596	3700	3345.1	86.13
	Curve	CRV	803	3493	3194.93	81.31
	Audius	AUDIO	1021	3275	3736.59	76.23

Table A.2: Daily Data Summary of Selected Cryptocurrencies Traded on Bitstamp

Type	Name	Symbol	No. of Observation	No. of zero volumes	No. of Non-zero volumes	Average volume	% Non-zero volumes
Native Coins							
	Bitcoin	BTC	364	0	364	2,289.05	100
	Stellar	XLM	364	0	364	6,882,900	100
	XRP	XRP	364	0	364	46,593,153	100
	Ethereum	ETH	364	0	364	16,487.85	100
	Litecoin	LTC	364	0	364	19,418.21	100
	Bitcoin Cash	BCH	364	0	364	5,184.64	100
	Cardano	ADA	364	0	364	469,651.9	100
	Algorand	ALGO	364	0	364	366,720.4	100
	Songbird	SGB	364	0	364	1,354,903	100
	Hedera Hashgraph	HBAR	364	0	364	715,269.8	100
Stable Coins							
	Tether	USDT	364	0	364	2,038,350	100
	USD Coin	USDC	364	0	364	146,280	100
	Dai	DAI	364	23	341	2,6096.82	93.68
	Paxos Standard Token	PAX	364	93	271	7,349.53	74.45
	Gemini Dollar	GUSD	364	247	117	1,185.03	32.14
	Tether EURt	EURT	364	286	78	374.92	21.43
Token							
	Chainlink	LINK	364	0	364	138,562.8	100
	Ploygon	MATIC	364	0	364	42,4164.5	100
	Uniswap	UNI	364	0	364	24,334.5	100
	AAVE	AAVE	364	0	364	1,169.16	100
	The Sandbox	SAND	364	0	364	75,935.64	100
	Curve	CRV	364	0	364	58,694.55	100
	Audius	AUDIO	364	0	364	95,783.09	100
	The Graph	GRT	364	0	364	241,152.1	100
	Gala	GALA	364	0	364	270,816.6	100
	Compound	COMP	364	0	364	467.91	100
	Chiliz	CHZ	364	0	364	11,472.7	100
	yearn.finance	YFI	364	1	363	4.12	99.73
	Maker	MKR	364	1	363	64.64	99.73
	Axie Infinity	AXS	364	1	363	826.11	99.73
	Basic Attention Token	BAT	364	1	363	6,6084.45	99.73
	Synthetix	SNX	364	3	361	7,58.75	99.18
	Enjin Coin	ENJ	364	7	357	17,460.33	98.08
	Smooth Love Potion	SLP	364	7	357	134,7507	98.08
	Kyber Network	KNC	364	9	355	14,314.1	97.53
	UMA	UMA	364	10	354	2,352.08	97.25
	Fantom	FTM	364	13	351	67,143.95	96.43
	dYdX	DYDX	364	17	347	3,393.68	95.33
	Fetch.ai	FET	364	28	336	34,016.57	92.31
	SKALE Network	SKL	364	35	329	101,611.7	90.38
	Storj	STORJ	364	55	309	4,560.35	84.89
	Alpha Finance	ALPHA	364	59	305	17,436.55	83.79
	Perpetual Protocol	PERP	364	117	247	1,431.04	67.86
	Shiba Inu	SHIB	364	156	208	2,725,982,421	57.14
	Swipe	SXP	364	220	144	425.39	39.56

Daily average volumes over the period 2022-01-01 to 2022-12-30 of 52 weeks in numbers of traded coins

Table A.3: Hourly PCA Eigenvalue, Variance Percentage Explained by Each PC and Cumulative Variance Percentage

	Eigenvalue	Variance Percent	Cumulative Variance Percent
Dim.1	13.36	44.55	44.55
Dim.2	1.29	4.31	48.86
Dim.3	1.26	4.19	53.05
Dim.4	1.05	3.52	56.57
Dim.5	1.02	3.41	59.98
Dim.6	0.98	3.27	63.25
Dim.7	0.87	2.91	66.16
Dim.8	0.84	2.81	68.97
Dim.9	0.74	2.47	71.44
Dim.10	0.72	2.39	73.83
Dim.11	0.65	2.15	75.98
Dim.12	0.62	2.07	78.04
Dim.13	0.59	1.95	80.00
Dim.14	0.58	1.93	81.92
Dim.15	0.53	1.75	83.68
Dim.16	0.52	1.73	85.40
Dim.17	0.49	1.65	87.05
Dim.18	0.47	1.58	88.63
Dim.19	0.44	1.46	90.09
Dim.20	0.39	1.30	91.40
Dim.21	0.37	1.22	92.62
Dim.22	0.36	1.19	93.81
Dim.23	0.34	1.13	94.93
Dim.24	0.31	1.04	95.97
Dim.25	0.28	0.95	96.92
Dim.26	0.27	0.89	97.81
Dim.27	0.22	0.73	98.54
Dim.28	0.18	0.60	99.13
Dim.29	0.17	0.57	99.70
Dim.30	0.09	0.30	100.00

The first principal component explains the largest percentage of total variance. The difference between the percentage of variance explained by the first and second principal components suggests a common factor in the cryptocurrency market.

Table A.4: Daily PCA Eigenvalue, Variance Percentage Explained by Each PC and Cumulative Variance Percentage

	Eigenvalue	Variance Percent	Cumulative Variance Percent
Dim.1	21.39	59.41	59.41
Dim.2	1.33	3.69	63.09
Dim.3	1.14	3.17	66.27
Dim.4	1.03	2.87	69.14
Dim.5	0.95	2.65	71.79
Dim.6	0.86	2.39	74.18
Dim.7	0.70	1.96	76.13
Dim.8	0.64	1.79	77.92
Dim.9	0.63	1.74	79.66
Dim.10	0.61	1.70	81.36
Dim.11	0.53	1.47	82.83
Dim.12	0.47	1.32	84.15
Dim.13	0.46	1.27	85.41
Dim.14	0.43	1.19	86.60
Dim.15	0.38	1.06	87.66
Dim.16	0.35	0.97	88.64
Dim.17	0.34	0.94	89.57
Dim.18	0.33	0.92	90.49
Dim.19	0.32	0.88	91.37
Dim.20	0.29	0.81	92.18
Dim.21	0.28	0.77	92.95
Dim.22	0.27	0.74	93.69
Dim.23	0.25	0.68	94.37
Dim.24	0.22	0.61	94.98
Dim.25	0.22	0.60	95.58
Dim.26	0.20	0.56	96.15
Dim.27	0.19	0.52	96.67
Dim.28	0.18	0.51	97.18
Dim.29	0.17	0.46	97.64
Dim.30	0.16	0.43	98.07
Dim.31	0.15	0.41	98.48
Dim.32	0.15	0.40	98.88
Dim.33	0.12	0.34	99.22
Dim.34	0.12	0.32	99.55
Dim.35	0.09	0.25	99.79
Dim.36	0.08	0.21	100.00

The first principal component explains the largest percentage of total variance. The difference between the percentage of variance explained by the first and second principal components suggests a common factor in the cryptocurrency market.

Table A.5: The Intercept of Hourly Functional Regression

Hour	Estimate	Std. Error	t value	P-value
1	-0.043	0.036	-1.216	0.227
2	0.038	0.041	0.941	0.349
3	-0.034	0.042	-0.814	0.417
4	0.030	0.030	1.015	0.313
5	-0.010	0.025	-0.386	0.700
6	-0.027	0.034	-0.783	0.435
7	0.018	0.028	0.636	0.526
8	0.016	0.028	0.557	0.579
9	0.030	0.025	1.196	0.235
10	-0.015	0.031	-0.476	0.635
11	0.010	0.028	0.365	0.716
12	-0.018	0.032	-0.567	0.572
13	0.054	0.037	1.455	0.149
14	0.001	0.041	0.031	0.975
15	0.004	0.040	0.100	0.921
16	-0.036	0.040	-0.889	0.376
17	0.034	0.038	0.892	0.375
18	-0.047	0.036	-1.310	0.193
19	0.095	0.036	2.645	0.009
20	0.038	0.043	0.893	0.374
21	0.039	0.027	1.431	0.155
22	0.043	0.041	1.060	0.292
23	-0.026	0.032	-0.810	0.420
24	-0.068	0.036	-1.907	0.059

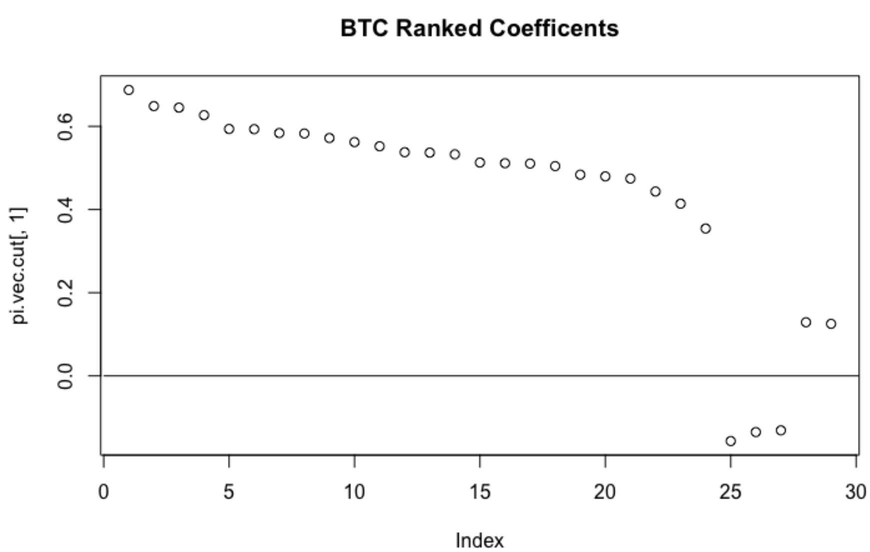
There is one statistically significant intercept at hour 19 which corresponds to the closing hour of NYSE

Table A.6: The Intercept of Daily Functional Regression

Day	Estimate	Std. Error	t value	P-value
Monday	0.262	0.332	0.788	0.436
Tuesday	0.023	0.208	0.113	0.911
Wednesday	0.190	0.269	0.708	0.484
Thursday	-0.031	0.233	-0.135	0.894
Friday	0.085	0.291	0.293	0.772
Saturday	-0.202	0.133	-1.514	0.139
Sunday	0.085	0.159	0.537	0.595

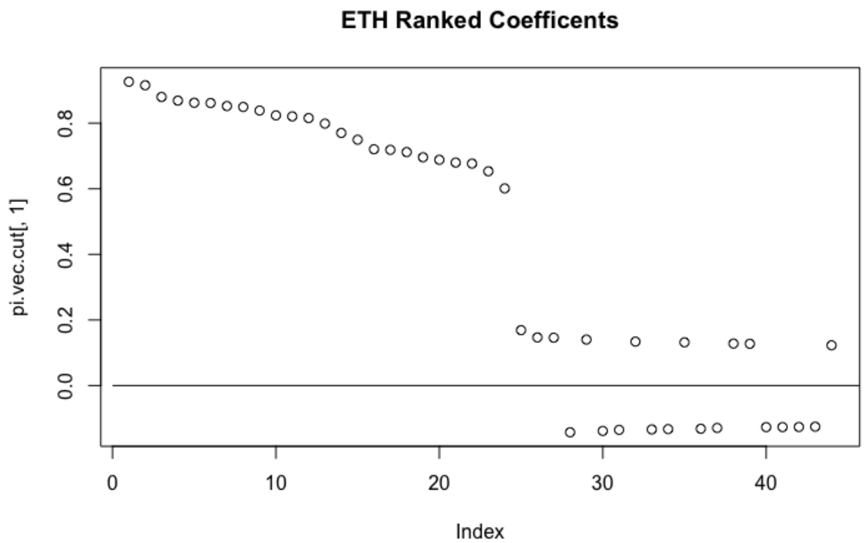
There is no statistically significant intercept at any day

Figure A.1: The Ordered Hourly Functional Regression Coefficients for BTC



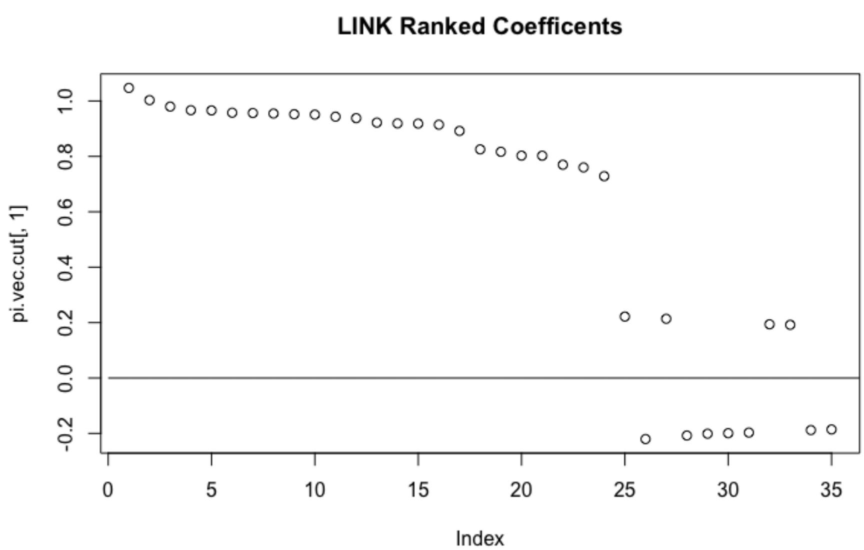
The coefficients are ordered by their absolute values

Figure A.2: The Ordered Hourly Functional Regression Coefficients for ETH



The coefficients are ordered by their absolute values

Figure A.3: The Ordered Hourly Functional Regression Coefficients for LINK



The coefficients are ordered by their absolute values

Table A.7: Description of Cryptocurrencies (sourced from Chatgdp and www. coinmarket-cap.com)

Bitcoin	BTC	It's the first and well-known decentralized digital currency used for peer-to-peer transactions and as a store of value..
Stellar	XLM	It focuses on facilitating fast and low-cost cross-border transactions and remittances. It aims to connect financial institutions and enable seamless money transfers.
XRP	XRP	It is designed for efficient cross-border payments and partnerships with financial institutions. It aims to improve international money transfers and settlement systems.
Ethereum	ETH	A versatile blockchain platform that allows developers to craft various smart contracts and decentralized applications (DApps).
Litecoin	LTC	operates on a decentralized network using blockchain technology but differs in its faster transaction times (processing blocks roughly every 2.5 minutes) and uses a different hashing algorithm called Scrypt. Its primary aim is to provide quicker transaction confirmations compared to Bitcoin.
Bitcoin Cash	BCH	A fork of Bitcoin, Bitcoin Cash strives to amplify scalability and transaction speed by enlarging the block size. It centers its efforts on serving as a digital currency suited for everyday spending.

Cardano	ADA	A blockchain platform recognized for its meticulous research-driven strategy and emphasis on security and scalability. It tries to establish a robust groundwork for constructing smart contracts and DApps.
Algorand	ALGO	A blockchain platform with high rapidity and scalability focuses on creating decentralized applications and provides a blockchain foundation with elevated processing capabilities.
Songbird	SGB	canary network for Flare, the blockchain built for universal connectivity.
Hedera Hashgraph	HBAR	is favoured by its high throughput and low-latency consensus. It targets applications that require both swift and secure transactional processes.
Tether	USDT	is one of the earliest and most well-known stablecoins. It's pegged 1:1 to the US Dollar and is designed to maintain a 1:1 value ratio.
USD Coin	USDC	is another popular stablecoin pegged 1:1 to the US Dollar. It's regulated and issued by several financial institutions, aiming to provide transparency and security.
Dai	DAI	A Ethereum blockchain decentralized stablecoin. It is generated through collateralized loans and algorithmic mechanisms to maintain its value at around 1 USD.

Paxos Standard Token	PAX	a stablecoin regulated by the New York State Department of Financial Services. It's pegged 1:1 to the US Dollar.
Gemini Dollar	GUSD	is issued by the Gemini cryptocurrency exchange. It is pegged 1:1 to the USD and fully backed by USD held in reserve accounts.
Tether EURt	EURT	EURT is a stablecoin pegged 1:1 to the Euro.
Chainlink	LINK	A decentralized oracle network that connects smart contracts on the blockchain with real-world data, events, and APIs. It aims to bridge the gap between blockchain and external data sources, enabling smart contracts to access and interact with information from the outside world in a secure and reliable manner.
Polygon	MATIC	Formerly known as Matic Network is a scaling solution for Ethereum that enhances its capabilities by providing a framework for building and connecting Ethereum-compatible blockchains. It seeks to address Ethereum's scalability issues and offers faster and cheaper transactions.

Uniswap	UNI	A decentralized exchange protocol built on the Ethereum blockchain. It enables users to trade Ethereum-based tokens directly from their wallets without the need for intermediaries. UNI is the governance token of the Uniswap protocol, allowing token holders to participate in decisions about its development and management.
AAVE	AAVE	A decentralized lending protocol built on the Ethereum blockchain. It allows users to lend, borrow, and earn interest on various cryptocurrencies in a peer-to-peer manner. AAVE is the native token of the Aave platform, used for governance and fee-sharing.
The Sandbox	SAND	A virtual world and gaming platform that operates on blockchain technology. It enables users to create, own, and monetize their gaming experiences and assets. SAND is the utility token used within The Sandbox ecosystem.
Curve	CRV	A decentralized exchange designed for stablecoin trading with low slippage. It focuses on providing efficient and low-risk trading of stablecoins like USDC, DAI, and USDT. CRV is the governance token of the Curve protocol.

Audius	AUDIO	A decentralized music-sharing and streaming platform that allows artists to directly share and monetize their music. AUDIO is the native token of the Audius platform, used for governance and rewarding network participants.
Synthetix	SNX	A decentralized synthetic asset platform built on Ethereum. It enables the creation and trading of synthetic assets that track the value of real-world assets like currencies, commodities, and cryptocurrencies. SNX is the token used to collateralize the creation of synthetic assets.
yearn.finance	YFI	A decentralized platform that aims to optimize yield farming strategies across different DeFi protocols. YFI is the governance token of the Yearn.finance ecosystem, allowing holders to participate in decisions about the platform's development.
The Graph	GRT	A decentralized indexing and query protocol for blockchain data. It allows developers to efficiently retrieve and index data from various blockchains, improving the functionality of decentralized applications. GRT is the native token used for protocol governance.

Gala	GALA	The native utility token of the Gala Games platform, which aims to empower game developers and players through blockchain technology. It allows players to participate in game development decisions and own in-game assets.
Kyber Network	KNC	A decentralized liquidity protocol that facilitates the exchange of tokens directly from wallets. It aims to provide a seamless and secure way to swap tokens without intermediaries.
Maker	MKR	The governance token of the MakerDAO platform, which is responsible for creating and managing the stablecoin DAI. MKR holders have the power to make decisions about the platform's parameters and upgrades.
Compound	COMP	The governance token of the Compound protocol, a decentralized lending platform. Users can lend and borrow cryptocurrencies, and COMP holders can propose and vote on changes to the platform.
Chiliz	CHZ	A cryptocurrency designed for sports and entertainment platforms, allowing fans to engage with their favorite teams through blockchain-based tokens and applications.
Axie Infinity	AXS	A blockchain-based game that enables players to collect, breed, and battle fantasy creatures known as Axies. AXS is the utility token used within the Axie Infinity ecosystem.

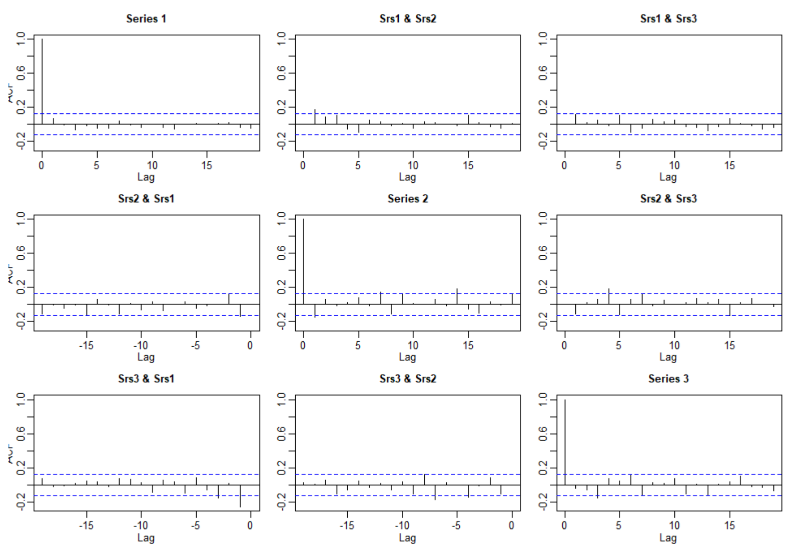
Basic Attention Token	BAT	The native token of the Brave browser ecosystem. It's designed to improve online advertising and content monetization by rewarding users for their attention and engagement.
Fantom	FTM	A high-performance blockchain platform that aims to supply fast and scalable smart contract capabilities. It's designed to address the limitations of existing blockchain networks.
UMA	UMA	A decentralized platform that allows users to create synthetic assets and financial contracts. It focuses on enabling decentralized finance (DeFi) innovations.
Enjin Coin	ENJ	Designed to facilitate the creation and management of non-fungible tokens (NFTs) for gaming and virtual goods. It aims to enable a thriving NFT ecosystem.
Smooth Love Potion	SLP	The in-game token used within the Axie Infinity game. Players earn SLP by participating in battles and breeding Axies.
Alpha Finance	ALPHA	The governance token of the Alpha Finance DeFi platform. It's used for voting on proposals and participating in the platform's development.
Fetch.ai	FET	An AI-driven blockchain platform that focuses on autonomous machine-to-machine communication and economic activity.

dYdX	DXDY	A decentralized exchange and DeFi platform that provides trading and lending services for various cryptocurrencies.
SKALE Network	SKL	A blockchain platform that offers high-performance and configurable sidechains to enhance Ethereum's scalability.
Storj	STORJ	A decentralized cloud storage platform that allows users to rent out their unused storage space and earn STORJ tokens in return.
Perpetual Protocol	PERP	A decentralized trading platform that delivers perpetual contracts for various cryptocurrencies.
Swipe	SXP	A multi-asset digital wallet and cryptocurrency payment platform aims to allow easy conversion between fiat currency and cryptocurrencies.
Shiba Inu	SHIB	Shiba Inu is a meme-inspired cryptocurrency that gained attention for its resemblance to other well-known cryptocurrencies. It's part of the broader "dogecoin" meme coin trend.

Appendix B

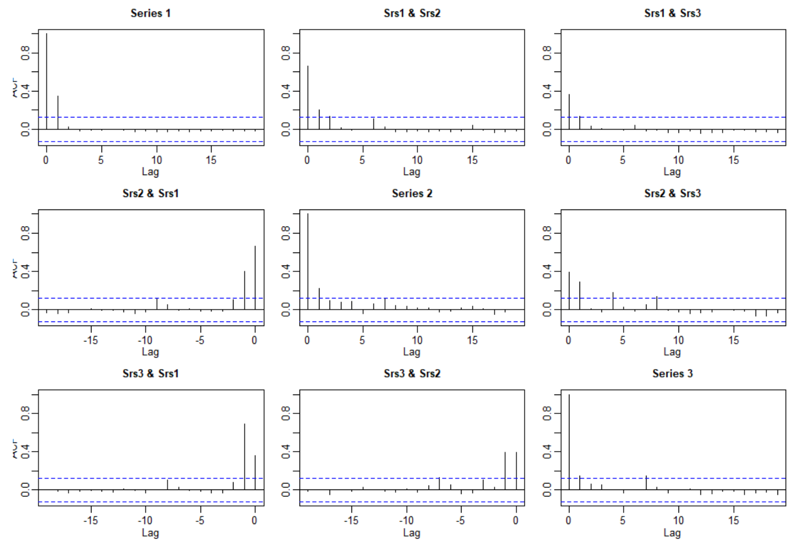
Technical Appendix for Chapter 4

Figure B.1: ACF of Hourly Eigenscores Corresponding with Eigenvalue Order 1 to 3



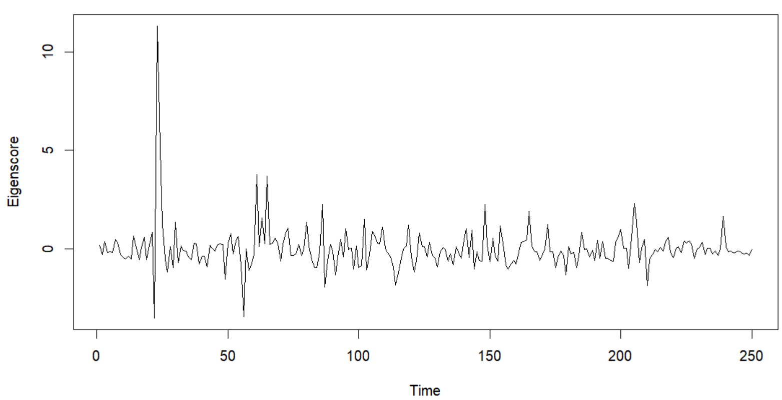
The diagonal graphs show the ACF of each eigenscore. The off-diagonal graphs show the cross ACF between different eigenscores.

Figure B.2: ACF of Hourly Square of Eigenscores Corresponding with Eigenvalue Order 1 to 3



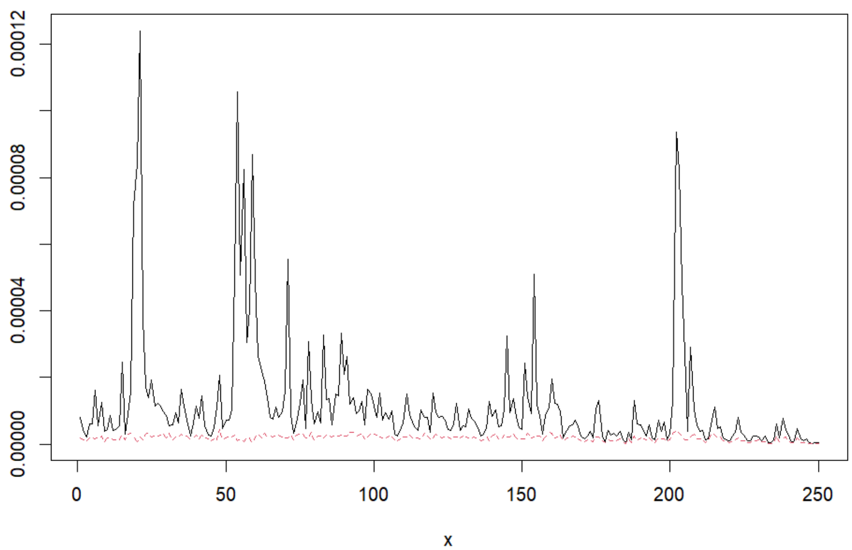
The diagonal graphs show the ACF of the square of each eigenscore. The off-diagonal graphs show the cross ACF between different squares of eigenscores.

Figure B.3: 1st Eigenscores Associated with the Largest Eigenvalue



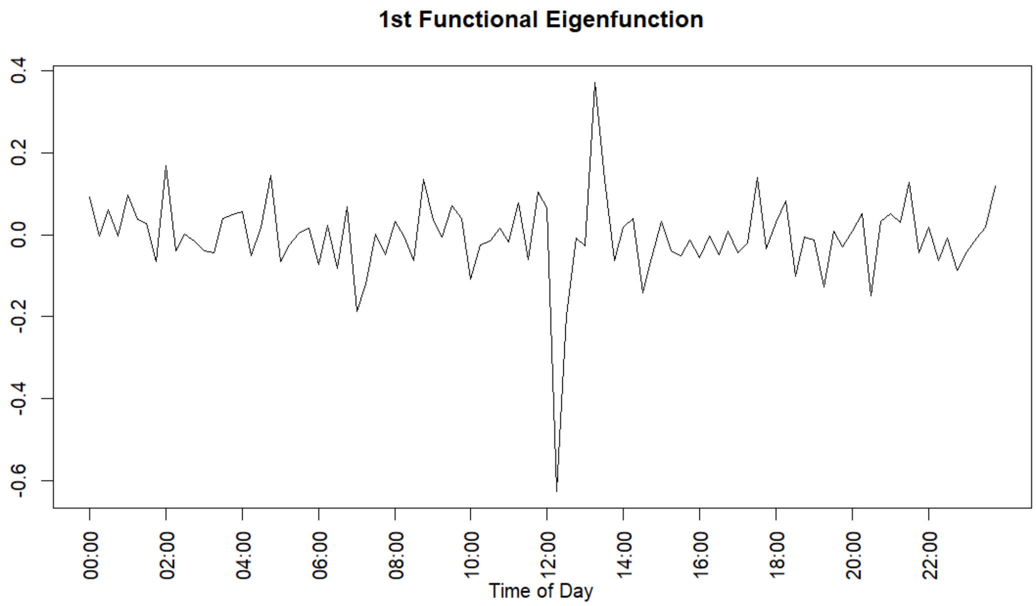
This figure shows the series of 250 first eigenscores associated with the largest eigenvalue.

Figure B.4: Daily Realized Variance of 15-Minute BTC Returns and Daily Variance of In-sample Estimation Errors



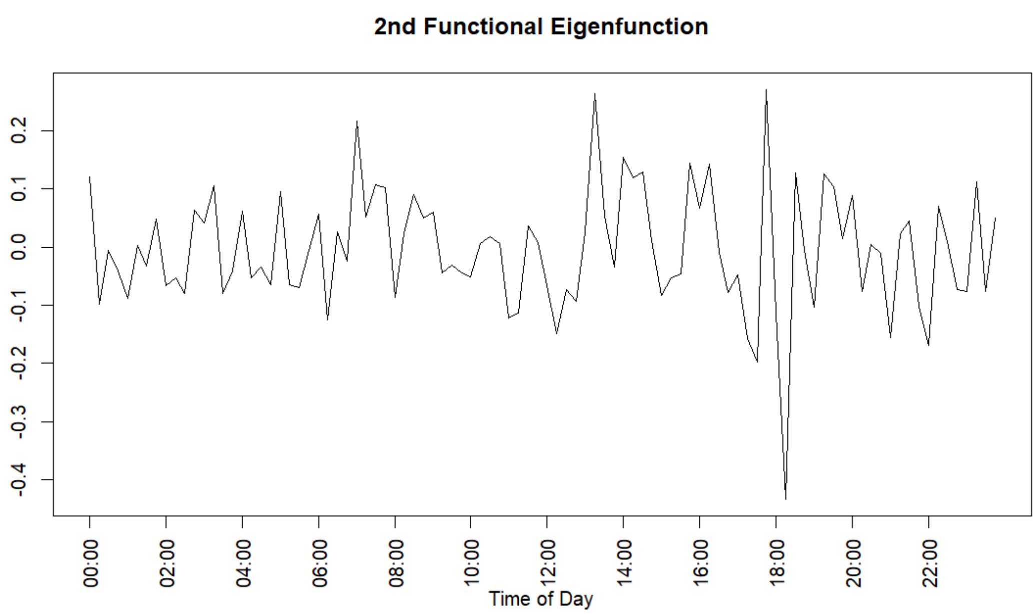
The black line illustrates the daily RVs of 15-Minute BTC returns times 10^{-4}, *where* $RV_i = \sum_{t=1}^{96} r_{i,t}^2$. *The red line illustrates the daily variance of the in-sample error* $\hat{e}_i(t) = X_i(t) - \hat{X}_i(t)$ *with a mean of* 1.95×10^{-6}.

Figure B.5: 1st Eigenfunction for 15-Minute BTC Returns



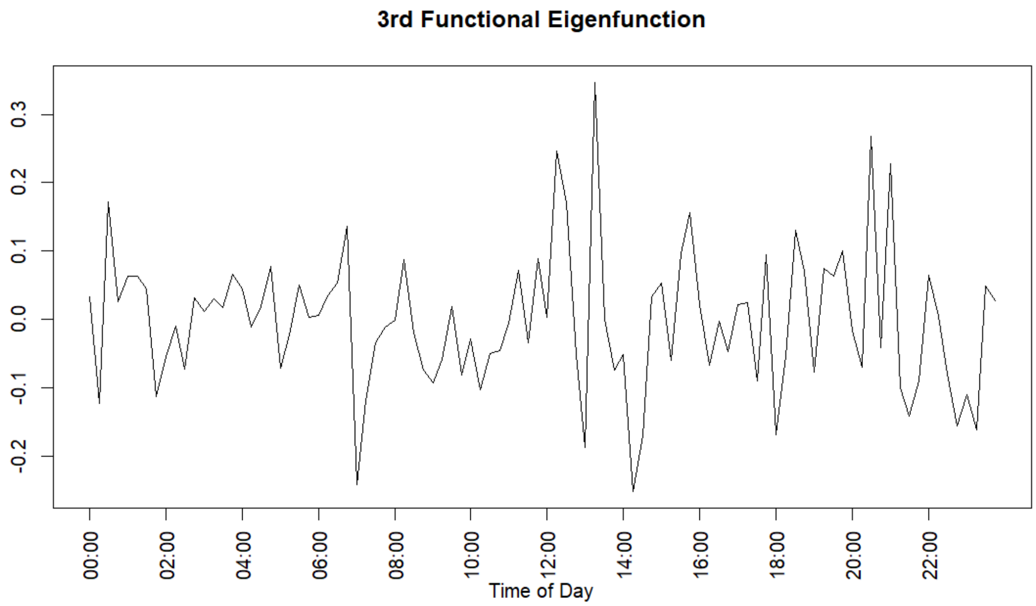
The UTC time is used here as the time reference

Figure B.6: 2nd Eigenfunction for 15-Minute BTC Returns



The UTC time is used here as the time reference

Figure B.7: 3rd Eigenfunction for 15-Minute BTC Returns



The UTC time is used here as the time reference

LM Test of the null hypothesis of the absence of conditional heteroscedasticity in returns.

Table B.1: BTC 15-Min Return LM Test Statistics at 5%

df (lag)	χ^2	p-value
1	36.32	< 0.001
2	38.50	< 0.001
3	43.37	< 0.001
4	43.74	< 0.001

The null hypothesis of the absence of conditional heteroscedasticity is rejected up to and including lag 4 at 5% significance level.

Table B.2: BTC Hourly Return LM Test Statistics at 5%

df (lag)	χ^2	p-value
1	18.26	< 0.001
2	18.34	< 0.001
3	18.51	< 0.001
4	20.33	< 0.001

The null hypothesis of the absence of conditional heteroscedasticity is rejected up to and including lag 4 at 5% significance level.

B.1 Constraints on the Parameters

This Appendix illustrates the constraints on the parameters implied by the marginal variances of scores.

a) univariate model

Let us suppose that the eigenscores $\beta_{i,j=1},..,\beta_{i,J_0}$ are independent and follow univariate AR(1)-GARCH$(1,1)$ models. Then, the eigenscore j on day i is:

$$\begin{aligned} \beta_i &= \mu + a\beta_{i-l} + \epsilon_i \\ \nu_i &= \varsigma_0 + \zeta\epsilon_{i-l}^2 + \varsigma\nu_{i-1}, \end{aligned} \tag{B.1.1}$$

where

$$\epsilon_i = \sqrt{v_i}z_i, \quad \forall i = 1,..,N \tag{B.1.2}$$

Then, the marginal variance σ^2 of ϵ_i satisfies:

$$\sigma^2 = \lambda_i - a^2 = \varsigma_0/(1 - (\zeta + \varsigma)) \tag{B.1.3}$$

where $\zeta + \varsigma \neq 1$ by the standard stationarity assumption on the GARCH(1,1).

b) multivariate model

$$\boldsymbol{\beta}_i = \mathbf{c} + \mathbf{\Pi}_1\boldsymbol{\beta}_{i-1} + \boldsymbol{\epsilon}_i \; i = 1, ...N \tag{B.1.4}$$

and the conditional variance of $\boldsymbol{\epsilon}_i$ is

$$\mathbf{H}_i = \mathbf{CC}' + a\boldsymbol{\epsilon}_{i-1}\boldsymbol{\epsilon}_{i-1}' + g\mathbf{H}_{i-1} \; i = 1, ...N \tag{B.1.5}$$

Then the marginal variance Σ of $\boldsymbol{\epsilon}_i$ satisfies

$$\Sigma = \Lambda - \Pi\Pi' = CC'/(\lambda_i - (a+g)) \tag{B.1.6}$$

where $\Lambda = Id[\lambda_1, ...\lambda_{J_0}]'$ and $a + g \neq 1$.

B.2 Basis Function Expansion

To transform functions into discrete-time sampling data we can use the basis function expansion. Suppose that $X_i(t)$ has mean 0, or has been demeaned. Then, each function X_i can be expressed as a linear combination of known basis functions ϕ_l, where $l = 1, ..., L$. For example, suppose that each function can be expressed by the basis expansion as

$$X_i(t) = \sum_{l=1}^{L} c_{i,l}\phi_l(t). \tag{B.2.1}$$

where, $X_i(t)$ is a 15-minute function, $t \in [1, \cdots, 96]$. In matrix terms:

$$X = C\phi \tag{B.2.2}$$

where the coefficient matrix C is $N \times L$ and the basis function ϕ is a $L \times T$ matrix where each row corresponding to the basis component $\phi_1, ..., \phi_L$. In matrix terms, the sample variance-covariance function is written as

$$cov(s,t) = N^{-1}\phi(s)'C'C\phi(t) \tag{B.2.3}$$

Define the symmetric matrix $W = \int \phi\phi'$ of dimension L with elements $w_{L_1,L_2} = \int \phi_{L_1}\phi_{L_2}$. If we choose the orthonormal basis function, such as the orthonormal Fourier series used in this paper, then $W = I$, the identity matrix of dimension L.

Following Ramsay and Silverman (2005b), we suppose that the eigenfunction ξ, solution of $\int cov(s,t)\xi(t)dt = \lambda\xi(t)$ for an appropriate eigenvalue λ has an expansion:

$$\xi(s) = \sum_{l=1}^{L} b_{i,l}\phi_l(s) = \phi(s)'b. \tag{B.2.4}$$

Then we get:

$$\int cov(s,t)\xi(t)dt = \int \frac{1}{N}\phi(s)'C'C\phi(t)'bdt = \phi(s)'\frac{1}{N}C'CWb \tag{B.2.5}$$

The eigenfunctions ξ_1 and ξ_2, for example, are orthogonal if the corresponding vectors of coefficients satisfy $\boldsymbol{b_1'Wb_2} = 0$. To get the required principal components, we define $\boldsymbol{u} = \boldsymbol{W^{0.5}b}$, solve the eigenvalue problem

$$\frac{1}{N}W^{0.5}C'CW^{0.5}u = \lambda u \tag{B.2.6}$$

Hence, to satisfy the orthogonalility condition, we compute $b = W^{-0.5}u$ for each eigenvector of that equation. If the basis is orthonormal, like the Fourier basis, then the FPCA problem is reduced to the standard multivariate PCA of $\boldsymbol{C}$, i.e., we only need to carry out the eigenanalysis of order L symmetric array $N^{-1}\boldsymbol{C'C}$ (Ramsay and Silverman, 2005b).

B.3 Pointwise Forecast Bands Algorithm

The pointwise forecast bands algorithm was introduced by Aue et al. (2015), and applied by Shang (2020) and Shang and Kearney (2022). It follows the steps given below. For ease of exposition, the horizon $h = 1$ is considered.

Step 1: Estimate J sample eigenscore vectors $(\hat{\beta}_1, \dots, \hat{\beta}_J)$ and sample eigenfunctions $[\hat{\xi}_1(t), \dots, \hat{\xi}_J(t)]$ using all N observations.

Step 2: For $k \in [J+1, \dots, N-1]$, calculate the in-sample errors of the forecast function

$$\hat{e}_{k+1}(t) = X_{k+1}(t) - \hat{X}_{k+1}(t), \tag{B.3.1}$$

where

$$\hat{X}_{k+1} = \sum_{j=1}^{J} \hat{\beta}_{(k+1)j} \hat{\xi}_j(t) \tag{B.3.2}$$

and $\hat{\beta}_{(k+1)j}$ are forecasted by the chosen univariate or multivariate time series model.

Step 3: For each time point t, define the standard deviation of the in-sample errors $\hat{e}_{k+1}(t)$ as:

$$\gamma(t) = \sqrt{\frac{\sum \hat{e}_{k+1}(t)^2}{(N-1) - (J+1)}}. \tag{B.3.3}$$

Step 4: Seek turning parameters $\underline{\kappa}_\alpha, \bar{\kappa}_\alpha$ such that $\alpha \times 100\%$ of the errors satisfy

$$-\underline{\kappa}_\alpha \gamma(t) \leq \hat{\epsilon}_k(t) \leq \bar{\kappa}_\alpha \gamma(t) \tag{B.3.4}$$

A practical approach to determine the pointwise forecast interval for a reasonable sample of size N, is:

$$\frac{1}{N-J-1}\sum_{k=1}^{N-J-1} I(-\underline{\kappa}_{\alpha}\gamma(t) \leq \hat{\epsilon}_k(t) \leq \bar{\kappa}_{\alpha}\gamma(t))$$
$$\approx P(-\underline{\kappa}_{\alpha}\gamma(t) \leq X_{k+1}(t) - \hat{X}_{k+1}(t) \leq \bar{\kappa}_{\alpha}\gamma(t)) \tag{B.3.5}$$

for all points of time $t = 1, \ldots, T$.