Master Thesis

# Edit: Master Thesis Title

Jorge Eduardo Frías Navarrete

Submitted in partial fulfillment of the requirement for the degree of:

Master of Science

| | |
|---|---|
| Student ID: | 012329686 |
| Degree programme: | Quantitative Finance |
| Supervisor: | Univ.Prof. David Preinerstorfer, Ph.D. |
| Date of Submission: | August 22, 2025 |

*Department of Finance, Accounting and Statistics.*
*Vienna University of Economics and Business.*
*Welthandelsplatz 1, 1020 Vienna, Austria.*

# Acknowledgements

Here write acknowledgements.

# Abstract

Here goes my abstract text. Here goes my abstract text. Here goes my abstract text. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse eu dolor luctus, rhoncus leo in, commodo turpis. Aenean sed enim in sem euismod porta. Vivamus tempor lorem nec eros rhoncus, eu hendrerit libero tincidunt. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque dapibus turpis quis nibh molestie dapibus. Aliquam erat volutpat. Integer et odio nec mauris sollicitudin mattis.

# Table of contents

# List of figures



# List of tables

# 1. Introduction

En alta volatilidad, quiza incluso mencionar los tweets de Elon Musk, por ejemplo, diciendo que Tesla no aceptaria pagos con bitcoin debido al alto costo energetico asociado con su mineria, causando una baja de los precios

Mirar Liu et al (y el otro paper similar) donde habla de las criptomonedas, e inspirarnos de ahi. Quiza la otra tesis en esto tambien.

Empezar con una introduccion de criptomonedas, del mercado, de la gran volatilidad, grandes retornos. Mencionar coin Market cap, la capitalizacion del mercado total de criptomonedas.

Mencionar articulos o reportes donde mencionen la importancia de este mercado, cuantas personas en promedio tienen criptos en su portafolio. Mencionar sucesos recientes importantes, como la introduccion de criptomonedas en algunos exchanges, de futuros en CME, de indices en XXX, del boom en la crisis de COVID (ver Mercik, donde menciona sucesos importantes).

Exhaustive list of predictors for the cross-section of stock returns Feng et al. (2020) A. Y. Chen & Zimmermann (2021)

mencionar a Ross, que comprobo la estructura lineal de los factores: – Esto quiza en introduccion

** Ejemplo de frase See for additional discussion of literate programming.**

## 1.1. Research objective

The main questions addressed in this thesis are:

- first research questions
- second research questions

Replicate the work of Kelly et al. (2019) and Bianchi & Babiak (2021) on the cryptocurrency market.

## 1.2. Literature review

Hacerla similar a la descripciond e la hipoteca inversa. Mencionar las dos corrientes de litertura: por una parte, los modelos de factores construidos como managed-strategies. Mencionar los modelos de factores mas conocidos, y alguna de la literatura importante en este tema: fama french modelo de tres factores Fama & French (1993), FF modelo de cinco factores Fama & French (2015) Citar la coleccion de Chen and ZImmerman A. Y. Chen & Zimmermann (2021) con la gran cantidad de factores en su dataset.

Por otra, una corriente basada en la estadistica que construye los factores mediante modelos puramente estadisticos, y que asume que los factores son latentes. Mencionar los dos autores pioneros de PCA: Chamberlain & Rothschild (1983) and Connor & Korajczyk (1986).

Mencionar modelos recientes con factores dinamicos. Describir brevemente estos modelos de PCA, y despues propuestas de modelacion de factores latentes dinamicos. Entre estos, hablar de Kelly et al. (2019), que propuso el modelo dinamico de factores, y mas recientemente, RPCA Q. Chen et al. (2022) Z. Chen et al. (2024), inspirado en regressiones de fama macbeth, que hace una combinacion de este modelo mas una implementacion de PCA.

Bianchi & Babiak (2021) aplicaron el modelo de Kelly en el mercado de criptos, mencionar otros papers que utilizaron el IPCA (hay otros de Kelly que lo uso para bonos y opciones, investigar)

Z. Chen et al. (2024) aplico el RPCA para el cross-section de diferentes asset-classes.

Mencionar literatura extensa tratando de entender los factores de riesgo de las criptomonedas, por ejemplo, Mercik et al. (2025), y otros papers que tengo en mis archivos y en notas.

Baur & Hoang (2021) Studies stablecoins as a safe heaven for Bitcoin volatility - safe heavens, but finds theyre not stable. Wang et al. (2020) also investigates if stablecoins are real diversifiers, hedgers or safe heavens. Hoang & Baur (2024) Analyses the stability of the so called stablecoins Asadov et al. (2023) Analyses precious-metal-backed coins as a better alternative for stablecoins.

## 1.3. Data concerns

One of the main challenges in this thesis was obtaining a large panel of cryptocurrency data. I extracted market data from the free CoinCodex API,which provides access to the full historical data of the cryptocurrencies listed on its platform. In contrast, most crypto market data providers –also called coin-ranking sites, such as CoinMarketCap, CoinGecko, CryptoCompare (CoinDesk)– offer limited access to historical data (usually one year) or none at all without a paid subscription. Some exchange platforms, such as Bybit, Binance, Coinbase, and Cex, allow users to extract market data for free through their public APIs. However, the number of cryptocurrencies (and thus, the cross-section) available from these sources was relatively small compared with CoinCoidex, and the time span was shorter [1].

The choice of which data source is appropriate for scientific research is subject to debate. For example, Alexander & Dakos (2020) examine different cryptocurrency data providers and find inconsistencies in regression estimates, suggesting that the source of cryptocurrency data can influence empirical results. Moreover, they document distorted coin prices on coin-ranking sites, caused by inflated or artificial trading volumes[2], emphasizing the importance of using traded data from crypto exchanges. By contrast, Vidal-Tomás (2022) argue that coin-ranking sites use the same underlying process as crypto exchanges and other platforms to compute a cryptocurrency price, and they report no significant differences in empirical results when using alternative data sources. To address these concerns, I apply a series of pre-processing filters, described in Chapter 3, to mitigate the impact of potential inaccuracies in my dataset.

[1] For example, bitcoin data started from late 2013 in CoinCodex, compared to November, 2022 in Bybit, January, 2019, in Binance, and June, 2021, in Coinbase. The available cryptocurrencies paired with Tether USD (USD) were 763 in Bybit, 623 in Binance, and 116 (USD) in Coinbase.

[2] Coin-ranking sites rank coins and exchanges by trading volume and market capitalization. As highlighted by Alexander & Dakos (2020), the prices quoted on some of these sites are calculated by aggregating the prices from hundreds of exchanges using a volume-weighted average. Because many exchanges artificially inflate their volume to boost their position in the rankings, the resulting aggregated prices are influenced by fake volumes and therefore inconsistent with traded prices.

# 2. Methodology

The Instrumented Principal Component Analysis (IPCA) model was introduced in the seminal work of Kelly et al. (2019, 2020). The main model used in this thesis is the IPCA with different $K$ number of factors

Explicar la metodologia de IPCA. Si hay tiempo, entonces explicar tambien como funciona el RPCA de Chen and Roussanov. Explicar las R2 (en lugra de R2, entonces poner Total score y predictive score), mencionar como pie de pagina que son las medidas definidas por Kelly et al. (2019).

Explicar los bootstrap para medir la significancia cada caracteristica, y quiza mencionar tambi'en brevemente los characteristic managed portfolios, en que consisten y como se emplean (quiza tomar inspiracion de Kelly, Bianchi, o creo que puede ser mejor en Liu et al.)

## 2.1. Instrumented Principal Component Analysis

# 3. Data

In this section, I introduce the cryptocurrency data used in this thesis, the series of filters applied to clean and prepare the dataset, and the summary statistics of the cryptocurrency excess returns. In addition, I show the set of asset-specific characteristics constructed from the cryptocurrency market data, which are used as instruments for latent factor exposures in the IPCA model. Appendix A.2 provides a detailed description of the characteristics used in the empirical analysis.

[++++ **ADD SMALL INTRO ABOVE OF RIK FACTORS CREATED** ++++]

The data extraction and pre-processing are primarily conducted in R 4.5.1 (R Core Team, 2025), using, among other packages[1], the `tidyverse` (v. 2.0.0; Wickham et al., 2019). Additional cleaning steps and visualizations are performed in Python 3.13.5 (Python Software Foundation, 2025). The full reproducible code is available in Appendix A.1.

## 3.1. Data extraction and sample construction

I collect daily cryptocurrency data on open, high, close, and low (OHCL) prices, 24-hour volume, and market capitalization (calculated as the cryptocurrency's USD price multiplied by its circulating supply) from CoinCodex, a website-data provider that gathers and aggregates data from more than 400 exchanges. I extract the data, all expressed in US dollars, using the CoinCodex API as follows:

1. I retrieve the list of all available cryptocurrencies and extract each cryptocurrency shortname, also referred to as the "slug". At the time of writing, there are 14,907 unique cryptocurrency shortnames listed in the API.

2. Using the slug, I construct an URL for each cryptocurrency to obtain the metadata from the API. I parse the JSON API response into a dataframe and extract

[1] See Appendix A.3 for the full list of software used in the empirical study.

the OHCL prices, volume, and market capitalization daily data. I exclude those observations with non-zero or missing values in any of these fields.

Out of the 14,907 cryptocurrencies listed, only 7,272 entries contained available data. Next, following the methodology of Bianchi & Babiak (2021) and Mercik et al. (2025), I apply a series of cleaning and filtering steps in order to remove possible innacuracies in the dataset:

1. Non-positive and missing values. As mentioned earlier, I remove observations where prices, volume, or market capitalization were non-positive or missing.

2. Small cryptocurrencies. Similar to Y. Liu et al. (2022), I screen out small cryptocurrencies and consider only those with a market capitalization greater than one million USD. Therefore, I exclude observations for coins whose market capitalization falls below this minimum threshold, which allows for the possibility that a coin may become "small" after a certain period or event.

3. Cryptocurrency type. Based on the cryptocurrency classification from CoinMarketCap and CoinCodex, I exclude:

   - stablecoins. I include (i) centralized stablecoins, which are backed and pegged to fiat currency or physical assets by a third party, such as Tether (USDT), USD Coin (USDC), and Euro Coin (EURC), and (ii) algorithmically stabilized stablecoins, which use algorithms to adjust the circulating supply in response to changes in demand to maintain a stable value with the underlying asset, such as DAI and AMPL (FSB, 2020).

   - wrapped cryptocurrency tokens, which mirror the value of another cryptocurrency from a different blockchain, e.g., Wrapped Bitcoin (wBTC) or Wrapped Ethereum (wETH) (Coinbase, n.d.).

   - cryptocurrencies backed by or pegged to gold or precious metals, including Pax Gold (PAXG) or XAGx Silver Token (XAGX).

4. Erroneous trading volume. To filter out cryptocurrencies with "fake" or "erroneous" trading volume, I calculate the daily volume-to-market-capitalization ratio for each token and exclude observations where the ratio exceeds 1.

5. Extreme returns. To minimize the influence of extreme values in my results, I winsorize daily cryptocurrency returns to lie within the range of -90% to 500%.

6. Time period. Even though cryptocurrency data are available since 2014, I use data from June 1, 2018 for the empirical analysis due to the low amount of coins

available before this date (see Figure 3.1).

7. Minimum observations. In order to maintain practical relevance, I keep cryptocurrencies that have at least 365 consecutive daily observations and those with at least 730 observations in the complete panel of coin characteristics (see Section 3.3), which is equivalent to 2 years of historical data. Therefore, I exclude very short-lived coins, but retain failed coins with this relatively large number of observations, which help to lessen the so called "survivorship biais".

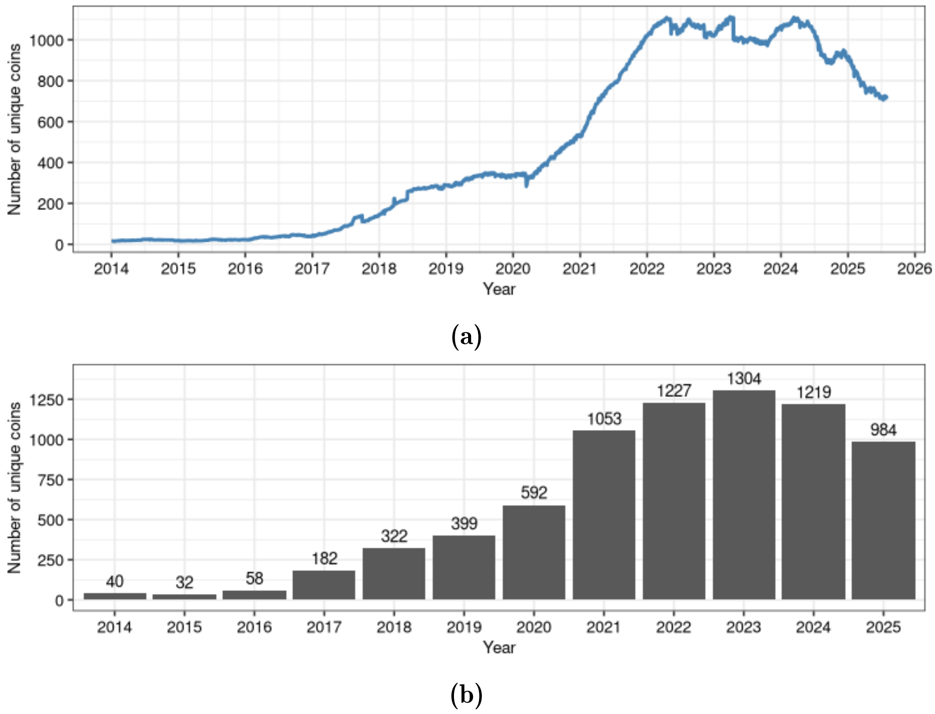


**Figure 3.1.: Number of cryptocurrencies over time.** Panel A shows the daily time series of the number of unique cryptocurrencies. Panel B displays the number of unique cryptocurrencies recorded each year. Both panels correspond to the dataset after applying the filtering steps (1) to (5), covering the period from January 1, 2014, to July 31, 2025, and including 1,416 unique cryptocurrencies. Note that coins may enter or exit the market over time.

## 3.2. Sample overview

After applying all the filters, the resulting sample consists of 973 unique cryptocurrencies and 1,478,936 observations from June 1, 2018, to July 31, 2025, where a day starts at 00:00:00 UTC. It is important to mention that the number of cryptocurrencies fluctuates over the entire period, which results in an unbalanced panel of data. Table 3.1 provides a description of the yearly cross-sectionional statistics: the sample starts with 254 different cryptocurrencies in 2018 and peaks in 2023 with 939 unique cryptocurrencies, before decreasing to 780 in 2025. The minimum daily cross-section is 121 in 2018,

**Table 3.1.: Cross-section size of the sample.** The table repots the number of unique coins per year, as well as the minimum daily cross-section size in the filtered sample.

| Year | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|---|---|---|
| Unique coins | 254 | 337 | 420 | 714 | 938 | 939 | 906 | 780 |
| Min. daily cross-section | 121 | 239 | 207 | 381 | 699 | 793 | 710 | 578 |

and then increases drastically up to 793 in 2023. For context, at the time of writing, CoinMarketCap tracks around 19 million cryptocurrencies, and CoinGecko around 19 thousands. When compared to these numbers, the size of the sample may seem small; however, it actually covers most of the whole cryptocurrency market capitalization (see Figure 3.2). The sample period includes important events in the market, such as

Table 3.2 summarizes the descriptive statistics for the cryptocurrency daily returns across different subsamples and Bitcoin, Ethereum, and Ripple, which are the three largest cryptocurrencies in the sample. Interestingly, the larger samples exhibit a larger volatility and more pronounced extreme returns, both positive and negative. Bitcoin shows the lowest mean return during the sample period (0.16% per day), though this value very close to that of Ethereum (0.17%) and Ripple (0.20%), and only slightly below other cryptocurrency subsamples.

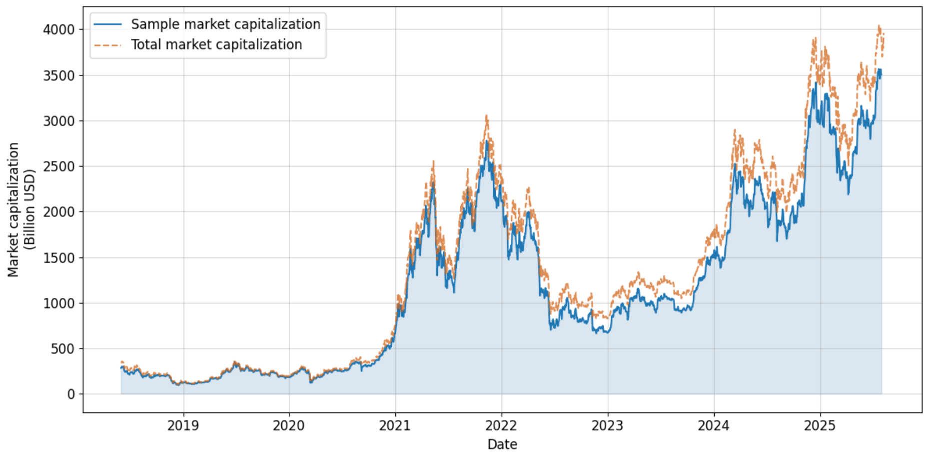


**Figure 3.2.: Cryptocurrency market capitalization**. The figure compares the cryptocurrency market capitalization in the filtered sample (blue line) with the total market capitalization (yellow line) from June 1, 2018 to July 31, 2025. Source: total market capitalization from CoinGecko.

The sample period spans several major market, economic, and political events, these include: the start of the COVID-19 pandemic and the subsequent crypto bubble in 2020-2021, El Salvador adoption of Bitcoin as legal tender in September 2021, and China's ban on cryptocurrency exchanges and mining in October 2021. The period also experienced multiple cryptocurrency exchange hacks[2], and geopolitical shocks such as

[2]For example, Binance, largest crypto exchange in the world, was hacked in 2019, and KuCoin and

**Table 3.2.: Summary statistics of daily returns.** The table reports summary statistics of daily returns for the filtered sample, the top 100 and top 10 cryptocurrencies ranked by market capitalization, and for Bitcoin, Ethereum, and Ripple individually. Reported statistics include the number of daily observations, the number of unique coins over the sample period, the mean and standard deviation of returns, and the 10th percentile, lower quartile, median, upper quartile, and 90th percentile of the distribution of the returns. The sample period is from June 1, 2018, to July 31, 2025.

| | No. Obs | Unique coins | Mean | Std | P10 | P25 | P50 | P75 | P90 |
|---|---|---|---|---|---|---|---|---|---|
| Sample | 1,478,936 | 973 | 0.36% | 12.25% | -6.83% | -3.00% | -0.16% | 2.57% | 6.85% |
| Top 100 | 176,400 | 100 | 0.21% | 6.93% | -5.64% | -2.52% | -0.03% | 2.44% | 5.86% |
| Top 10 | 24,747 | 10 | 0.25% | 5.74% | -4.71% | -2.00% | 0.07% | 2.14% | 5.07% |
| Bitcoin | 2,618 | 1 | 0.16% | 3.33% | -3.24% | -1.27% | 0.09% | 1.52% | 3.67% |
| Ethereum | 2,611 | 1 | 0.17% | 4.35% | -4.33% | -1.77% | 0.10% | 2.14% | 4.88% |
| Ripple | 2,540 | 1 | 0.20% | 5.31% | -4.48% | -1.87% | 0.08% | 1.89% | 4.70% |

[1] As of July 31, 2025, the top 10 cryptocurrencies are Bitcoin, Ethereum, Ripple, Binance Coin, Solana, Dogecoin, Tron, Cardano, Stellar, and Chainlink.

the Russia-Ukraine war in February 2022, and the Palestine-Israel war in October 2023. More recently, in 2024, the U.S. Secutities and Exchange Commission (SEC) approved the listing and trading of several crypto spot ETFs in January, and Donald Trump's election as U.S. president, with Elon Musk playing and important role in his campaign (Bianchi & Babiak, 2021; C. Chen & Liu, 2022; S. Liu & Yang, 2024; Mercik et al., 2025; Zhou, 2025).

## 3.3. Characteristic construction and description

For the analysis, I construct 41 asset-specific characteristics from the cross-section of 973 cryptocurrencies using data on prices, volume, and market capitalization. Specifically, I follow the methodology of Bianchi & Babiak (2021), Y. Liu et al. (2022), and Mercik et al. (2025) to replicate and classify a set of characteristics that serve as return predictors in the empirical analysis. Based on economic intuition, I group the characteristics into six categories: market and size, volatility and risk, trading activity, liquidity, past returns, and distribution. Table 3.3 summarizes the set of characteristics, while Appendix A.2 provides a detailed description and the construction procedure.

This is more related to factor construction.

Organize week in the following way: the first seven days of the year forms the first week, and the first 51 weeks of the year consists of 7 days each. The 52th week of the year consists of the last eight days and, in case of a leap year (as 2016, 2020, and 2024), of nine days.

Crypto.com were hacked in 2020 and 2022, respectively. (Zhou, 2025)

**Table 3.3.: Cryptocurrency characteristics.** The table presents the 41 cryptocurrency characteristics used as return predictors in the empirical analysis. The characteristics are grouped in six categories: price and size, volatility and risk, trading activity, liquidity, past returns, and distribution.

| No. | Characteristic | Symbol | Definition |
|---|---|---|---|
| **Panel A: Price & size** | | | |
| (1) | Market capitalization | mcap | Last day's market capitalization. |
| (2) | Price | prc | Last day's logged closing price. |
| (3) | Closeness to the 90-day high | dh90 | Last day's price over the maximum price in the previous 90 days. |
| **Panel B: Volatility & risk** | | | |
| (4) | Market beta | beta | CAPM market beta, estimated from 30 days of daily returns. |
| (5) | Idiosyncratic volatility | ivol | Volatility of CAPM residuals over 30 days of daily returns. |
| (6-7) | Realized volatility | rvol_*d | Realized volatility, calculated from 7 and 30 days of OHCL prices. |
| (8) | Return volatility | retvol | Standard deviation of daily returns over 7 days. |
| (9) | Value-at-Risk | var | The historical Value-at-Risk at 5% level over 90 days. |
| (10) | Expected Shortfall | es_5 | The expected shortfall at the 5% level over 90 days. |
| (11) | Price delay | delay | Improvement in $R^2$ after adding lagged one-and two-day market excecss return to the CAPM. |
| **Panel C: Trading activity** | | | |
| (12) | Trading volume | volume | Last day's daily trading volume. |
| (13) | Average volume | volume_*d | Mean volume over the past 7 and 30 days. |
| (15) | Turnover | turn | The last day's trading volume over current market capitalization. |
| (16) | Average 7-day turnover | turn_7d | Mean turnover over the past 7 days. |
| (17) | Turnover volatility | std_turn | Turnover volatility over the past 30 days. |
| (18) | Trading volume volatility | std_vol | Volume's logged volatility over the past 30 days. |
| (19) | Volume's coefficient of variation | cv_vol | Volume's volatility over its mean in the previous 30 days. |
| **Panel D: Liquidity** | | | |
| (20) | Bid-ask spread | bidask | Mean estimated bid-ask spread calculated over the past 30 days. |
| (21) | Illiquidity | illiq | Mean absolute daily return over trading volume over the past 30 days. |
| (22) | Standardized abnormal turnover | sat | Last day's turnover minus its 30-day average, divided its volatility over 30 days. |
| (23) | De-trended turnover | dto | De-trended turnover minus the value-weighted daily market turnover. |
| (24) | Volume Shock 15-day | volsh_15d | Log deviation of trading volume from its rolling 15-day average. |
| (25) | Voume Shock 30-day | volsh_30d | Log deviation of trading volume from its rolling 30-day average. |
| **Panel E: Past returns** | | | |
| (26) | Daily reversal | r2_1 | Return on the previous trading day. |
| (27-30) | Momentum | r*_1 | 7, 14, 21, and 30-day cumulative return ending 1 day before the prediction date. |
| (31) | Intermediate momentum | r30_14 | Cumulative return from 30 to 14 days before the prediction date. |
| (32) | Long-term reversal | r180_60 | Cumulative return from 180 to 60 days before the prediction date. |
| (33) | CAPM alpha | alpha | CAPM intercept, estimated from 30 days of daily returns. |
| **Panel F: Distribution** | | | |
| (34-35) | Skewness | skew_*d | Skewness of the daily return distribution over a 7-and 30-day period. |
| (36-37) | Kurtosis | kurt_*d | Kurtosis of the daily return distribution over a 7-and 30-day period. |
| (38-39) | Maximum daily return | maxret_*d | The maximum daily return in the past 7-and 30 days. |
| (40-41) | Minimum daily return | minret_*d | The minimum daily return in the past 7-and 30 days. |

Following Y. Liu et al. (2022), I construct a daily cryptocurrency market return as the value-weighted average return of all the cryptocurrencies in the sample. For cryptocurrencies $i = 1, ..., N$, the daily market return at time $t$ is computed as:

$$r_t^M = \frac{\sum_{i=1}^{N} r_{it} \cdot marketcap_{it}}{\sum_{i=1}^{N} marketcap_{it}}$$

The cryptocurrency market excess return is constructed as the difference between the cryptocurrency market return and the risk-free rate. To proxy the risk-free rate, I used the (daily) 1-month Treasury bill rate from the FRED.

Write this in the following section of "Empirical application" or This is for the model: 7. (Still undecisive) Minimum cross-section. Following the criterion by Kelly, I Convert variables in the -0.5 - 0.5 range

The sample period ranges from January 1st, 2014, to May 31st, 2025.

# 4. Results

Implemented in python, based on the IPCA python code of Seth Pruitt [1] and the `ipca` python package of Buechner & Bybee (2019) [2].

Following Kelly et al. (2019), I cross-sectionally transform the instrument variables period-by-period in the following manner: first,

**Important**: mention the shift of characteristics: the conditional APT of Kelly, Pruitt, Su (JFE 2019) says that the characteristics known at Date=d-1 determine the exposures associated with the returns realized at Date=d; hence, here we should have shifted the characteristics in Z relative to the returns in R

This is a template of the table of the results of the IPCA model. I need to add a caption to the table. Here I reference Table 4.1.

Some test for quarto and latex

Quarto: 1. Sees the caption line after the table. 2. Wraps the `tabular` inside a LaTeX `table` environment. 3. Adds `\caption{Some letters with LaTeX}` and `\label{tbl-letters}` automatically. 4. Gives it a table number and puts it in the List of Tables.

---

**Inline LaTeX way inside Quarto**
Here we see the summary statistics in Table 4.2.

[1] See https://sethpruitt.net/research/.
[2] See https://bkelly-lab.github.io/ipca/.

**Table 4.1.: Results of IPCA regression.** Model Performance. Panel A and B report total and predictive $R^2$ in percent for the restricted ($\Gamma_\alpha = 0$) and unrestricted ($\Gamma_\alpha \neq 0$) IPCA model for $K$ number of factors on daily and weekly data, respectively. Panel C reports the corresponding total and predictive $R^2$ for a simple PCA model on weekly data.

| | | K | | |
|---|---|---|---|---|
| | | $K = 3$ | $K = 5$ | $K = 8$ |
| Panel C: PCA on weekly data | | | | |
| $R^2_{\text{Total}}$ | | 0.0000 | 0.0000 | 0.0000 |
| $R^2_{\text{Predictive}}$ | | 0.0000 | 0.0000 | 0.0000 |
| Panel B: IPCA on weekly data | | | | |
| $R^2_{\text{Total}}$ | $\Gamma_\alpha = 0$ | 0.2625 | 0.2817 | 0.2934 |
| | $\Gamma_\alpha \neq 0$ | 0.2661 | 0.2826 | 0.2937 |
| $R^2_{\text{Predictive}}$ | $\Gamma_\alpha = 0$ | 0.1725 | 0.1551 | 0.1511 |
| | $\Gamma_\alpha \neq 0$ | 0.1719 | 0.1584 | 0.1554 |
| Panel A: IPCA on daily data | | | | |
| $R^2_{\text{Total}}$ | $\Gamma_\alpha = 0$ | 0.2301 | 0.2509 | 0.2681 |
| | $\Gamma_\alpha \neq 0$ | 0.2322 | 0.2524 | 0.2690 |
| $R^2_{\text{Predictive}}$ | $\Gamma_\alpha = 0$ | -0.3904 | -0.4082 | -0.4169 |
| | $\Gamma_\alpha \neq 0$ | -0.3857 | -0.4055 | -0.4156 |

**Table 4.2.:** Some letters with LaTeX

| | | |
|---|---|---|
| A | B | C |
| D | E | F |

# 5. Conclusion

# References


Alexander, C., & Dakos, M. (2020). A critical investigation of cryptocurrency data and analysis. *Quantitative Finance*, *20*(2), 173–188. https://doi.org/10.1080/14697688.2019.1641347

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., Dervieux, C., & Woodhull, G. (2025). *Quarto* (Version 1.7) [Computer software]. https://doi.org/10.5281/zenodo.5960048

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56. https://doi.org/10.1016/S1386-4181(01)00024-6

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, *61*(1), 259–299. https://doi.org/10.1111/j.1540-6261.2006.00836.x

Ardia, D., Guidotti, E., & Kroencke, T. A. (2024). Efficient estimation of bid–ask spreads from open, high, low, and close prices. *Journal of Financial Economics*, *161*, 103916. https://doi.org/10.1016/j.jfineco.2024.103916

Asadov, A., Yildirim, R., & Masih, M. (2023). Toward greater stability in stablecoins: Empirical evidence from an analysis of precious metals. *Borsa Istanbul Review*, *23*(5), 1152–1172. https://doi.org/10.1016/j.bir.2023.07.004

Babiak, M., & Erdis, M. B. (2022). *Variations in Trading Activity, Costly Arbitrage, and Cryptocurrency Returns* (SSRN Scholarly Paper 4291073). Social Science Research Network. https://doi.org/10.2139/ssrn.4291073

Baur, D. G., & Hoang, L. T. (2021). A crypto safe haven against Bitcoin. *Finance Research Letters*, *38*, 101431. https://doi.org/10.1016/j.frl.2020.101431

Bianchi, D., & Babiak, M. (2021). *Mispricing and Risk Compensation in Cryptocurrency Returns* (SSRN Scholarly Paper 3935934). Social Science Research Network. https://doi.org/10.2139/ssrn.3935934

Buechner, M., & Bybee, L. (2019). *ipca: Instrumented principal components analysis* [Computer software]. https://github.com/bkelly-lab/ipca

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, *51*(5), 1281–1304. https:

//doi.org/10.2307/1912275

Chen, A. Y., & Zimmermann, T. (2021). *Open Source Cross-Sectional Asset Pricing* (SSRN Scholarly Paper 3604626). Social Science Research Network. https://doi.org/10.2139/ssrn.3604626

Chen, C., & Liu, L. (2022). How effective is China's cryptocurrency trading ban? *Finance Research Letters*, *46*, 102429. https://doi.org/10.1016/j.frl.2021.102429

Chen, Q., Roussanov, N. L., & Wang, X. (2022). *Semiparametric Conditional Factor Models in Asset Pricing* (SSRN Scholarly Paper 3984633). Social Science Research Network. https://doi.org/10.2139/ssrn.3984633

Chen, Z., Roussanov, N. L., Wang, X., & Zou, D. (2024). *Common Risk Factors in the Returns on Stocks, Bonds (and Options), Redux* (SSRN Scholarly Paper 4703281). Social Science Research Network. https://doi.org/10.2139/ssrn.4703281

Coinbase. (n.d.). *What is wrapped crypto?* Retrieved August 6, 2025, from https://www.coinbase.com/learn/your-crypto/what-is-wrapped-crypto

Connor, G., & Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics*, *15*(3), 373–394. https://doi.org/10.1016/0304-405X(86)90027-9

Datar, V. T., Y. Naik, N., & Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, *1*(2), 203–219. https://doi.org/10.1016/S1386-4181(97)00004-9

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, *75*(3), 1327–1370. https://doi.org/10.1111/jofi.12883

Financial Stability Board. (2020). *Addressing the regulatory, supervisory and oversight challenges raised by "global stablecoin" arrangements: Consultative document*. https://www.fsb.org/2020/04/addressing-the-regulatory-supervisory-and-oversight-challenges-raised-by-global-stablecoin-arrangements-consultative-document/

Garfinkel, J. A. (2009). Measuring Investors' Opinion Divergence. *Journal of Accounting Research*, *47*(5), 1317–1348. https://doi.org/10.1111/j.1475-679X.2009.00344.x

Garfinkel, J. A., Hsiao, L., & Hu, D. (n.d.). Disagreement and returns: The case of cryptocurrencies. *Financial Management*, *n/a*(n/a). https://doi.org/10.1111/fima.12491

George, T. J., & Hwang, C.-Y. (2004). The 52-Week High and Momentum Investing. *The Journal of Finance*, *59*(5), 2145–2176. https://doi.org/10.1111/j.1540-6261.2004.00695.x

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hoang, L. T., & Baur, D. G. (2024). How stable are stablecoins? *The European Journal of Finance*, *30*(16), 1984–2000. https://doi.org/10.1080/1351847X.2021.1949369

Hou, K., & Moskowitz, T. J. (2005). Market Frictions, Price Delay, and the Cross-Section of Expected Returns. *The Review of Financial Studies*, *18*(3), 981–1020. https://doi.org/10.1093/rfs/hhi023

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524. https://doi.org/10.1016/j.jfineco.2019.05.001

Kelly, B. T., Pruitt, S., & Su, Y. (2020). *Instrumented Principal Component Analysis* (SSRN Scholarly Paper 2983919). Social Science Research Network. https://doi.org/10.2139/ssrn.2983919

Komsta, L., & Novomestky, F. (2022). *moments: Moments, cumulants, skewness, kurtosis and related tests*. https://doi.org/10.32614/CRAN.package.moments

Lewellen, J., & Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, *82*(2), 289–314. https://doi.org/10.1016/j.jfineco.2005.05.012

Liu, S., & Yang, C. (2024). Spot cryptocurrency ETFs: Crypto investment products or stepping stones toward tokenization. *Finance Research Letters*, *69*, 106150. https://doi.org/10.1016/j.frl.2024.106150

Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common Risk Factors in Cryptocurrency. *The Journal of Finance*, *77*(2), 1133–1177. https://doi.org/10.1111/jofi.13119

Llorente, G., Michaely, R., Saar, G., & Wang, J. (2002). Dynamic Volume-Return Relation of Individual Stocks. *The Review of Financial Studies*, *15*(4), 1005–1047. https://doi.org/10.1093/rfs/15.4.1005

Mercik, A., Bdowska-Sójka, B., Karim, S., & Zaremba, A. (2025). Cross-sectional interactions in cryptocurrency returns. *International Review of Financial Analysis*, *97*, 103809. https://doi.org/10.1016/j.irfa.2024.103809

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250. https://doi.org/10.1016/j.jfineco.2011.11.003

Peterson, B. G., & Carl, P. (2024). *PerformanceAnalytics: Econometric tools for performance and risk analysis*. https://doi.org/10.32614/CRAN.package.PerformanceAnalytics

Posit team. (2025). *RStudio: Integrated development environment for r*. Posit Software, PBC. http://www.posit.co/

Python Software Foundation. (2025). *Python programming language* (Version 3.13.5) [Computer software]. https://www.python.org/

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ryan, J. A., & Ulrich, J. M. (2025). *quantmod: Quantitative financial modelling framework*. https://doi.org/10.32614/CRAN.package.quantmod

Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods – a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, *23*, 1164–1167.

The pandas development team. (2020). *Pandas-dev/pandas: pandas* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.3509134

Vaughan, D. (2024). *slider: Sliding window functions*. https://doi.org/10.32614/CRAN.package.slider

Vidal-Tomás, D. (2022). Which cryptocurrency data sources should scholars use? *International Review of Financial Analysis*, *81*, 102061. https://doi.org/10.1016/j.irfa.2022.102061

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, G.-J., Ma, X., & Wu, H. (2020). Are stablecoins truly diversifiers, hedges, or safe havens against traditional cryptocurrencies as their name suggests? *Research in International Business and Finance*, *54*, 101225. https://doi.org/10.1016/j.ribaf.2020.101225

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., …

Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Zeileis, A., & Grothendieck, G. (2005). zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, *14*(6), 1–27. https://doi.org/10.18637/jss.v014.i06

Zhou, F. (2025). Application of Event Study Methodology in the Analysis of Cryptocurrency Returns. *Emerging Markets Finance and Trade*, *61*(4), 989–1009. https://doi.org/10.1080/1540496X.2024.2404173

# A. Appendix

## A.1. Supplementary Material

## A.2. Cryptocurrency Characteristics

Following Bianchi & Babiak (2021), Y. Liu et al. (2022), and Mercik et al. (2025), I construct 41 asset-specific characteristics from OHCL prices, volume, and market capitalization of each cryptocurrency and group them into six categories: prize and size, volatility and risk, trading activity, liquidity, past returns, and distribution. The following list provides the definition of each characteristic and a description of their construction.

#### Price and size

**`mcap`**. Last day's market capitalization. The market capitalization is the current cryptocurrency circulating supply multiplied by its current price in USD.

**`prc`**. Last day's logged closing price.

**`dh90`**. The closeness to the 90-day high is defined as the ratio of the last day's price to the maximum price observed over the past 90 days (e.g., George & Hwang, 2004).

#### Volatility and risk

**`beta`**. The market beta is calculated as the slope coefficient from a 30-day rolling regression of cryptocurrency's excess returns on the market portfolio excess returns (e.g., Lewellen & Nagel, 2006). The coin market portfolio is constructed daily as the value-weighted average of cryptocurrency returns in the sample.

**`ivol`**. Idiosyncratic volatility is computed as the standard deviation of the residuals from the 30-day rolling CAPM regression, following the same approach as for `beta`.

**rvol_*d**. Realized volatility, computed using the estimator of Yang and Zhang (2000) based on OHCL prices. I compute the daily realized volatility over rolling 7-and 30-day windows, denoted rvol_7d and rvol_30d, respectively. For $n > 1$ number of periods, the volatility estimate at time $t$ is:

$$\sigma_t = \sqrt{\sigma_O^2 + k\sigma_C^2 + (1-k)\sigma_{RS}^2}$$

where $\sigma_{RS}^2$ is the variance estimator of Rogers et al. (1994), and $\sigma_O^2$, $\sigma_C^2$, $k$ are given by

$$\sigma_O^2 = \frac{1}{n-1}\sum_{i=1}^{n}(o_i - \bar{o})^2,$$

$$\sigma_C^2 = \frac{1}{n-1}\sum_{i=1}^{n}(c_i - \bar{c})^2,$$

$$k = \frac{\alpha - 1}{\alpha + \frac{n+1}{n-1}}$$

with $o = ln\, O_t - ln\, C_{t-1}$, and $c = ln\, C_t - ln\, O_t$. Here, $C_{t-1}$ denotes the previous day's closing price and $O_t$ the current day's opening price. I set the constant $\alpha = 1.34$, following Yang and Zhang (2000), who recommend this as the best value in practice.

**retvol**. Standard deviation of daily returns over the past 7 days (e.g., Ang et al., 2006).

**var**. The historical Value-at-Risk at the 5% level, based on daily returns over the past 90 days.

**es_5**. The expected shortfall at the 5% level, based on daily returns over the past 90 days.

**delay**. From the regression

$$R_i - R_f = \alpha^i + \beta^i_{CMKT}CMKT + \beta^i_{CMKT_{-1}}CMKT_{-1} + \beta^i_{CMKT_{-2}}CMKT_{-2} + \epsilon_i,$$

where $R_i$ is the return on asset $i$, $R_f$ is the risk-free rate, and $CMKT$, $CMKT_{-1}$, and $CMKT_{-2}$ are the current, lagged one-and two-day coin market portfolio excess returns, delay is the improvement in $R^2$ relative to the standard CAPM regression using only the current market portfolio excess returns (e.g., Hou & Moskowitz, 2005). The coin market portfolio is constructed as in beta.

**Trading activity**

**volume**. Last day's daily trading volume. The trading volume is the total amount of a cryptocurrency exchanged in a given day, measured in USD.

**volume_*d**. The average trading volume over the past 7 and 30 days, denoted volume_7d and volume_30d, respectively.

**turn**. Turnover, computed as the last day's trading volume over the current market capitalization (e.g., Datar et al., 1998).

**turn_7d**. Average turnover over the past 7 days.

**std_turn**. The standard deviation of the turnover over the past 30 days.

**std_vol**. The log standard deviation of trading volume over the past 30 days.

**cv_vol**. The coefficient of variation is the standard deviation of the daily trading volume divided by its mean, over the past 30 days (e.g., Babiak & Erdis, 2022).

**Liquidity**

**bidask**. The cryptocurrency bid-ask spread, computed from OHCL prices using the approximation of Ardia et al. (2024).

**illiq**. The Amihud (2002) price impact (illiquidity) measure, computed as the 90-day average of the ratio of the absolute daily return to daily trading volume.

**sat**. The standardized abnormal turnover, following Garfinkel et al. (n.d.). The measure is calculated as the last day's turnover minus its average over the past 30 days, divided by the turnover's standard deviation over the same 30-day period.

**dto**. De-trended turnover (e.g., Garfinkel, 2009). It is computed as turnover minus the value-weighted average daily market turnover, de-trended by its 180-day median.

**volsh_*d**. Volume shock, defined as the log-deviation of daily trading volume from its $k$-day rolling average (e.g., Llorente et al., 2002). For volsh_15d and volsh_30d, $k = 15$ and $k = 30$, respectively. For cryptocurrency $i$ at time $t$:

$$v_{i,t} = \log(\text{Volume}_{i,t}) - \log\left(\frac{1}{k}\sum_{s=1}^{k}\text{Volume}_{i,t-s}\right)$$

### A.2.1. Past returns

**r2_1**. Daily reversal, defined as the previous day's cryptocurrency return.

**r*_1**. The 7, 14, 21, and 30-day momentum, denoted `r7_1`, `r14_1`, `r21_1`, and `r30_1`, respectively. Momentum is defined as the cumulative return from the previous $k \in \{7, 14, 21, 30\}$ days up to one day before the return prediction.

Moskowitz et al. (2012)

**r30_14**. Cumulative return from the previous 30 days up to 14 days before the return prediction.

**r180_60**. Cumulative return from the previous 180 days up to 60 days before the return prediction.

**alpha**. The CAPM alpha, defined as the intercept from a 30-day rolling regression of cryptocurrency's excess returns on the market portfolio excess returns. The market portfolio is constructed as in `beta`.

### A.2.2. Distribution

**skew_*d**. Skewness of daily returns over the previous 7 and 30 days, denoted `skew_7d` and `skew_30d`, respectively.

**kurt_*d**. Kurtosis of daily returns over the previous 7 and 30 days, denoted `kurt_7d` and `kurt_30d`, respectively.

**maxret_*d**. The maximum daily return over the past 7 and 30 days, denoted `maxret_7d` and `maxret_30d`, respectively.

**minret_*d**. The minimum daily return over the past 7 and 30 days, denoted `minret_7d` and `minret_30d`, respectively.

## A.3. Software

This thesis was fully written using Quarto (Allaire et al., 2025), running in RStudio (v. 2025.5.1.513; Posit team, 2025) on Fedora Linux 42 (Workstation Edition).

I used R 4.5.1 (R Core Team, 2025) and the following R packages: bidask v. 2.1.4 (Ardia et al., 2024), moments v. 0.14.1 (Komsta & Novomestky, 2022), pcaMethods v. 2.0.0 (Stacklies et al., 2007), PerformanceAnalytics v. 2.0.8 (Peterson & Carl, 2024), quantmod v. 0.4.28 (Ryan & Ulrich, 2025), slider v. 0.3.2 (Vaughan, 2024), tidyverse v. 2.0.0 (Wickham et al., 2019), and zoo v. 1.8.14 (Zeileis & Grothendieck, 2005).

Additionally, I used Python 3.15.3 (Python Software Foundation, 2025) and the following packages: numpy (Harris et al., 2020), pandas (The pandas development team, 2020), matplotlib (Hunter, 2007), and scipy (Virtanen et al., 2020).