MASTERARBEIT / MASTERS THESIS

Master Thesis Title

verfasst von / submitted by

Jorge Eduardo Frías Navarrete

Submitted in total fulfillment of the requirement of the degree of:

Master of Science (WU)

Matrikelnummer / student ID number:	012329686
Studium / degree programme:	Quantitative Finance
University:	Vienna University of Economics and Business
Betreut von / Supervisor:	Univ.Prof. David Preinerstorfer, Ph.D.

Wien, August 2025 / Vienna, August 2025

Table of contents

List of Figures

List of Tables

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

```
1 + 1
```

```
[1] 2
```

1 Introduction

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).

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

** Ejemplo de frase See (**knuth84?**) for additional discussion of literate programming.**

```
1 + 1
```

```
[1] 2
```

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.

3 Methodology

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.)

4 Data

4.1 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 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 with where prices, volume, or market capitalization were non-positive or missing.

2. Small cryptocurrencies. Similar to 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

Table 4.1: **Summary statistics of daily excess returns.**

No. Obs	Unique coins	Min No. Obs	Mean	Std	P10	P25	P50	P75	P90
1478936	973	121	-2.70%	12.49%	-10.18%	-6.65%	-3.70%	0.02%	4.69%

Some text for the footnote

demand to maintain a stable value with the underlying asset, such as DAI and AMPL (FSB, 2020).

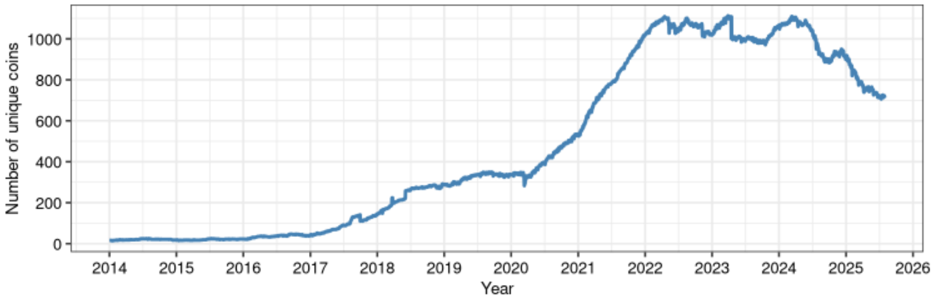
- wrapped cryptocurrency tokens, which mirror the value of another cryptocurrency from a different blockchain, e.g., Wrapped Bitcoin (wBTC) or Wrapped Ethereum (wETH) (Coinbase, n.d.).
- cryptocurrencies backed by or pegged to gold or precious metals, including Pax Gold (PAXG) or XAGx Silver Token (XAGX).

4. Volume-to-market-capitalization ratio. 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 4.1).
7. 6. 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 4.3), 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".

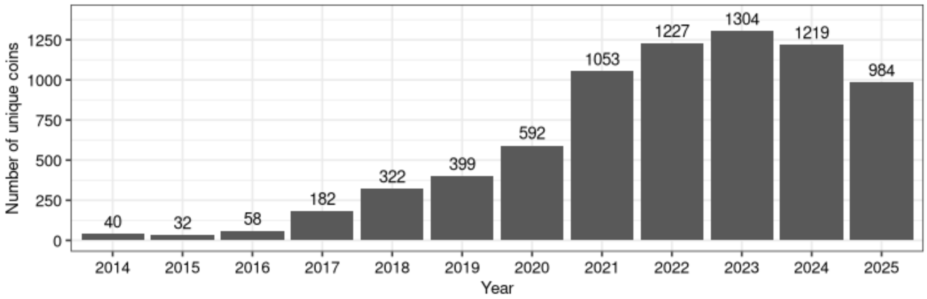
4.2 Sample overview

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 4.1 provides an overview of the descriptive statistics for the cryptocurrency excess returns. Additionally,

SECOND TRY



(a)



(b)

Figure 4.1: 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.

No. Obs	Unique coins	Min No. Obs	Mean	Std	P10	P25	P50
1478936	973	121	-0.027021	0.124855	-0.101770	-0.066473	-0.037020

TEST

```r
tab_latex <-
  gtcars |>
  dplyr::select(mfr, model, msrp) |>
  dplyr::slice(1:5) |>
  gt() |>
  tab_header(
    title = md("Data listing from **gtcars**"),
    subtitle = md("`gtcars` is an R dataset")
  ) |>
  as_latex()

tab_latex
```

ANOTHER TEST

THIS WORKS

		K		
		$K = 3$	$K = 5$	$K = 8$
Panel C: PCA on weekly data				
R^2_{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_{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_{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

Data listing from **gtcars**
`gtcars` is an R dataset

mfr	model	msrp
Ford	GT	447000
Ferrari	458 Speciale	291744
Ferrari	458 Spider	263553
Ferrari	458 Italia	233509
Ferrari	488 GTB	245400

```
library(knitr)
library(kableExtra)
df <- data.frame(v1=rnorm(6), v2=runif(6), v3=rbinom(6, 1, .33),
             row.names=LETTERS[1:6])

df$v4 <- c('My formula $\\sum_{i=1}^9$')
kable(df, "latex", align="c", booktabs=TRUE, escape = F, caption = '$\\Gamma$') %>%
footnote(general=c("$a^2+b^2=c^2,$",
                   "$\\\\sigma^2=\\\\frac{1}{n-1}\\\\sum_{i=1}^n(x_i-\\\\bar{x})^2;$",
                   "1,000 \\\\$;", "100\\\\%."),
```

Table 4.2: Γ

...1	No. Obs	Unique coins	Min No. Obs	Mean	Std	P10	P25
Sample	1478936	973	121	-0.0270213	0.1248547	-0.1017704	-0.0664732

Note: $a^2 + b^2 = c^2$, $\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$; 1,000 $; 100%.
[1] Hello there! *Hello there!*

Table 4.3: Γ

	v1	v2	v3	v4
A	-0.5862756	0.2395745	0	My formula $\sum_{i=1}^{9}$
B	-0.8392342	0.4948087	0	My formula $\sum_{i=1}^{9}$
C	0.4308115	0.9745896	0	My formula $\sum_{i=1}^{9}$
D	0.4799528	0.9035687	0	My formula $\sum_{i=1}^{9}$
E	1.1666455	0.1880054	1	My formula $\sum_{i=1}^{9}$
F	-0.2399438	0.6749801	0	My formula $\sum_{i=1}^{9}$

Note: $a^2 + b^2 = c^2$, $\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$; 1,000 $; 100%.

[1] Hello there! *Hello there!*

```
number=c("Hello\ there! \\\\textit{Hello\ there!}"),
footnote_as_chunk=TRUE,
escape=FALSE)
```

4.3 Characteristic construction and description

4.3.1 Volume shock

Following Bianchi et al. (2022), the volume shock is defined as the log-deviation of trading volume from its rolling average (over 30 or 60 days) for cryptocurrency i at time t. For $m \in \{30, 60\}$ periods, the volume shock is estimated as:

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

4.4 Risk

4.4.1 Realized volatility (rvol)

Using the volatility estimator of Yang and Zhang (2000), I compute the daily realized volatility based on OHCL prices over a rolling 30-day window. 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 σ_{RS}^2 is the variance estimator of Rogers et al. (1994), and σ_O^2, σ_C^2, k are defined as follows:

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

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

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

with $o = ln\,O_t - ln\,C_{t-1}$, and $c = ln\,C_t - ln\,O_t$. Here, C_{t-1} denotes the last days' closing price and O_t the current day's opening price. I set the constant $\alpha = 1.34$ as suggested by Yang and Zhang (2000) to be the best value in practice.

Moskowitz et al. (2012)

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.

Similar to **Liu et al**, 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 (`CMKT`) 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.

References

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

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

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, 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

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

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/
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
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
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
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
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