Universidad de los Andes
School of Engineering and Applied Sciences

Formulation of an asset pricing model for cryptocurrencies

Luca Appiani Caro

Dissertation for the degree of
Civil Industrial Engineer

Guiding teacher: Javier Mella

ING-IN-001/11

Santiago, June 2024

Certifico que he leído esta memoria y que en mi opinión su alcance y calidad son completamente adecuados como para ser considerada una memoria conducente al título de Ingeniero.

Javier Mella
(Profesor Guía)

Certifico que he leído esta memoria y que en mi opinión su alcance y calidad son completamente adecuados como para ser considerada una memoria conducente al título de Ingeniero.

Nombre Representante del Decano

Certifico que he leído esta memoria y que en mi opinión su alcance y calidad son completamente adecuados como para ser considerada una memoria conducente al título de Ingeniero.

Nombre Profesor Invitado

Summary

Acknowledgements

I dedicate this work to ...

Contents

Summary iv

Acknowledgements v

1 Introduction **2**

2 Theoretical Framework **4**

2.1 Literature review 4

2.1.1 Empirical Studies 4

2.1.2 Theoretic Models 9

2.1.3 Initial proposal of factors 10

3 Objectives and Methodological Framework **12**

3.1 Objectives 12

3.1.1 General Objective 12

3.1.2 Specific Objectives 12

3.2 Methodological Framework 13

3.2.1 Data Retrieval 13

3.2.2 Mathematical Formulation . . . 14
3.2.3 Fama Mac-Beth Regressions . . . 15
3.2.4 Gibbons, Ross, and Shanken Test . . . 16
3.2.5 Conclusions and recommendations . . . 17

4 Methodological Development **18**

4.1 Mathematical Formulation . . . 18
4.1.1 Math Detail . . . 19
4.2 Data . . . 25
4.2.1 Market Index . . . 26
4.2.2 Return computation . . . 26
4.2.3 "Popular" Factor Estimation . . . 27
4.3 Portfolio Building . . . 29
4.4 Fama Mac-Beth Regressions . . . 30

5 Results **31**

References **32**

Appendix **35**

A Technical details, tables, and others **35**

A.1 Covariance Definition . . . 35

A.2 Stein's Lemma 35

A.3 Popular Portfolio 35

List of Figures

4.1 Histogram of portfolio returns for “popular” and “not popular” cryptocurrencies. 28

List of Tables

4.1 T-Test Results. 28

4.2 Mean Portfolio Returns. 28

A.1 List of Popular Cryptocurrencies and their Symbols 36

Todo list

I need to establish: Type of study, Period and place where the research was done, Universe and sample, Methods, Variable selection, Assumptions, Procedures, Methods of data collection. 18

Add source. 26

Chapter 1

Introduction

The cryptocurrency world is a very intriguing one, the high volatility and technologies the assets that form part of this market support, provoke a lot of interest. The fact this type of assets are decentralized, have a lack of regulatory oversight, and operate on a global scale, pose significant challenges for investors and financial institutions.

Nonetheless, there has been a surge in investment options related to this asset class. Said growth is driven by several factors which include: the increasing demand from investors for exposure to cryptocurrencies, the emergence of new technologies, and the growing interest from institutional investors. It is crucial to highlight that this popularity extends not only to the asset class as a whole but also to the cryptocurrencies themselves.

With the increasing demand for effective models that can analyze and forecast the returns of cryptocurrencies, the popularity of these digital assets also continues to grow. Current empirical models in the literature offer valuable insights, yet frequently have a base of solid theoretical basis. Integrating finance and economics theories into cryptocurrency modeling can help grasp the factors influencing cryptocurrency returns and enhance the precision of prediction models.

This research plans to fill the gap in the literature by creating a model based on Portfolio Markowitz theory to improve the analysis of cryptocurrency returns. This model can offer investors and financial institutions important information on how to build portfolios, manage risks, and make investment decisions in the quickly changing cryptocurrency market. In the end, the goal of this study is to improve the comprehension of how cryptocurrency returns work and set the stage for smarter investment decisions in this developing asset category.

Chapter 2

Theoretical Framework

2.1 Literature review

Due to the rising popularity in recent years of cryptocurrency, there has been much research related to digital currency, from which the field of asset pricing is no exception. The latter because there is a growing interest related to the study of the factors that affect the returns of this type of assets, which certainly translates into a lot of studies whose objective is the previously mentioned. While the research topics may seem similar, it is important to note that this allows for a comprehensive categorization of the studies, despite the broadness of the related research.

2.1.1 Empirical Studies

The first group corresponds to the empirical studies that test for the performance of widely accepted asset pricing models such as the CAPM (Sharpe, 1964, Lintner, 1965 and Mossin, 1966), FF3 (Fama & French, 1993), FF5 (Fama & French, 2015), (Carhart, 1997), among others. The methodology is based on the recollection of data related to returns on a specific set of cryptocurrencies in a particular period to calculate the factors of the models mentioned previously. Due to the significant amount of investigation that follows said framework, there are also many studies that, in addition to the steps mentioned previously, complement with techniques that help understand better the underlying phenomena. For a better grasp of these groups of studies, some will be discussed that will most definitely aid the current investigation.

The first study included in the group of empirical studies is the one done by (Gregoriou, 2019). In this investigation they demonstrate that investors obtain abnormal excess returns on the London Stock Exchange from 2014 to 2017. The main reason behind this was because of earlier studies, like (Bariviera, 2017), that found evidence of inefficiency and lack of regulation related to the cryptocurrency market. The data used corresponds to daily returns of all London Stock Exchange listed securities from the years 2014-2017, where they conclude that by applying CAPM, FF3, Carhart, and FF5, investors do indeed obtain excess returns by speculating in cryptocurrencies, suggesting that they are inefficient. While this dissertation primarily does not explore into the efficiency of cryptocurrency markets, the insights from (Gregoriou, 2019) underline the broad applicability and versatility of such studies.

Another study is the one done by (Liu et al., 2022), where they find that there are three factors that capture the cross-sectional expected cryptocurrency returns. Despite not forming part of the core of the investigation, this study mentions a relevant aspect corresponding to the different opinions people have related to cryptocurrency; they say there are two views about the related market. The first one says that all coins represent bubbles and fraud. On the other hand, the second states technology behind said markets may become an important innovation and that at least some coins may become assets that represent a stake in the future of the related technology (Liu et al., 2022).

With the current information of cryptocurrency markets it is difficult to establish right from wrong with respect to those opinions. However, either way, empirical studies like (Liu et al., 2022) contribute largely to understand the factors that better explain the returns of corresponding assets. Regarding the research itself, the factors studied were cryptocurrency size, momentum, volume, and volatility. It is important to mention that the study focuses only on those market-factors, because financial and accounting data[1] was not available for the cross-section of the coins that were analyzed in the data.

Regarding the conclusions drawn from (Liu et al., 2022), there are several to con-

[1] Referring to information related to a company's performance, revenue, expenses, financial statements, which are crucial for assessing a company's financial position.

sider. Firstly, size and momentum factors well capture the cross-section of cryptocurrency returns. Furthermore, a three-factor model can be constructed using market information that is successful in pricing the strategies in the cryptocurrency market. A number of theoretical explanations are drawn for the factors. In relation to the cryptocurrency size premium, which refers to the phenomenon where the average returns of small firms are higher than those of large firms (Song, 2023). Said effect can be applied to cryptocurrencies. The cryptocurrency size factor relates to the liquidity effect, which encompasses the ease, speed, and affordability that an investor can trade a certain asset (Hasan et al., 2022). Secondly, they find some evidence that the size premium is consistent with a mechanism proposed by cryptocurrency theories[2]: the trade-off between capital gains and the convenience yield[3].

As to momentum, the conclusions show that they are in line with the investor overreaction channel, indicating the tendency of investors to react disproportionately to new information, which in turn causes the price of cryptocurrency to swing more than it should according to its intrinsic value (Diaconaşu et al., 2022).

Continuing the line of empirical validation studies, (Thoma, 2020) investigated whether an investing strategy modeled by Cumulative Prospect Theory (CPT) leads to a risk-adjusted outperformance, based on different factor models which include the (Fama & French, 1993). Cumulative Prospect theory is a model proposed by (Kahneman & Tversky, 1979), fits well in modeling how investors inform themselves about a certain cryptocurrency since they usually look at the price chart and then mentally represent a historical return distribution.

So, according to (Thoma, 2020), by looking at the price chart of cryptocurrency, investors evaluate the skewness[4] and evaluate the asset as a gamble, similar to lottery. The conclusions imply that cryptocurrency holders choose high prospect theory values over low values[5], with investors generally favoring the latter. Due to this predilection, cryptocurrencies with high prospect theory values are overbought, which reduces future gains. Cryptocurrencies with low prospect theory values on the other hand, are less likely to be overbought and might result in larger future returns. While the

[2]Some of them include (Sockin & Xiong, 2020), (Prat et al., 2019), and (Cong et al., 2020).
[3]According to (Hull, 2012) it corresponds to the benefits from holding the physical asset.
[4]Measure of the asymmetry of a distribution.
[5]It is important to note that high or low prospect theory values refers to the ones that are obtained in the expression used to calculate the respective factor used in the study, that is implemented in the factor models used in the investigation.

previous study presented did not place significant emphasis on the regression models themselves, it uses those models to complement the main model of the investigation, which was the prospect theory model.

Another approach to the asset pricing of cryptocurrency is the one taken by (Hayes, 2017), where a regression model is estimated using cost of production factors, rather than the usual market factors that comprise the most popular asset pricing models. Concerning the factors, (Hayes, 2017) concludes that more than 84% of value formation can be explained by three variables: computational power (as a representative for mining difficulty), rate of coin production, and the relative hardness of the mining algorithm employed.

(Shen et al., 2020) also followed a similar framework to the ones already mentioned. This study proposes a three factor pricing model, consisting of market, size and reversal factors. This model is compared with cryptocurrency-CAPM or C-CAPM, which uses only excess market returns to explain returns of cryptocurrency portfolios. As to the conclusions, the three-factor model based on the three factors already mentioned, has a better performance than the C-CAPM at explaining the cryptocurrency returns.

The research carried out by (Erfanian et al., 2022) provides another way of looking at the asset pricing of cryptocurrencies, while maintaining, to some extent, the empirical studies framework mentioned in the beginning of this literature review. They apply a series of machine learning approaches to investigate whether macroeconomic, microeconomic, technical, and blockchain indicators based on economic theories can predict bitcoin prices. Regarding the factor-based conclusions, based on a multilinear regression, the most significant long-term predictors were those of a macroeconomic nature, as well as blockchain information. Moreover, the empirical results showed that SVR (Support Vector Regressions) is the best machine learning model, and the effectiveness of feature selection techniques varied, with no clear winner emerging. Thus, providing evidence of the superiority of machine learning models in comparison to traditional methods for Bitcoin price prediction that use empirical research.

(Grobys & Sapkota, 2019) investigate about the popular momentum strategy implemented in the cryptocurrency market. Although there is no use of the more popular asset pricing models mentioned in the beginning (Sharpe, 1964), (Fama & French, 1993), (Fama & French, 2015), and (Carhart, 1997), in this case a time series approach is taken, that uses the return of a security over the past months to determine the investor position on said security in the following month. They do not find any significant evidence as to relevant momentum payoffs in the cryptocurrency market.

The research done by (Cai & Zhao, 2024) uses salience theory of choice under risk to show that investor behavior drives cross-sectional cryptocurrency returns. The reason being that headlines have significantly influenced the crypto asset class, sparking investor fear of missing out on the "crypto rush". A salience payoff refers to a payoff that stands out from the average, which under the context of salience theory, draws the attention of the investor. To examine salience payoffs (Cai & Zhao, 2024) construct a salience measure, which measures the difference between salience and equally weighted returns during a specific time period, weekly or monthly. To construct the ST measure, they follow the study of (Cosemans & Frehen, 2021). The empirical study of (Cai & Zhao, 2024) contains two parts; the study of the predictability of ST on cross-sectional crypto returns, and the investigation of the viability of ST as a cross-sectional pricing factor.

They conclude that given the asset class lacks fundamentals and has a concentrated clientele, the ST effect documented in the study is the strongest in the literature. In addition, they mention that ST is much more relevant for emerging assets that have high uncertainties. However, as the crypto market becomes more mainstream and attracts more institutional investors, the ST effect may lose its relevancy in explaining the return dynamics in the crypto market.

Due to the extensive amount of research related to the empirical studies, a final investigation will be presented, but it is important to mention that there are much more variants of this type of studies. (Long et al., 2020) research the cross-sectional seasonality anomaly in cryptocurrency markets. Said anomaly suggests that assets with highest (lowest) average same-calendar month return tend to overperform (underperform) in the future. In simpler terms, if an investor plans to invest on a Monday, she or he should check which assets delivered the highest returns on Mondays in the past. The models used in this case include CAPM and FF3. As to the conclusions, results demonstrate that there is a strong and sizable seasonality phenomena. However, they emphasize a limitation of their study relating the short sample period.

2.1.2 Theoretic Models

Now, concerning the second category of studies, they correspond to theoretic models or models that are derived from a theoretical framework. It is important to note that, unlike the empirical validation research, the quantity of theoretical models is much less. Particularly, studies focusing on cryptocurrencies are notably scarce. In despite of said shortage, one related study was found.

(Koutmos & Payne, 2021) developed an intertemporal regime-switching asset pricing model characterized by heterogeneous agents that have different expectations in relation to the volatility of the prices of bitcoin. The fact that models are intertemporal, refers to the fact that the models take into account changes in market conditions and risks over time; and as to the regime-switching part, this means that said models can switch between different states or "regimes", that could represent market conditions. Regarding the agents, there are three: mean-variance optimizers, speculators, and fundamentalists. Although the derivation of the model in this research does not come from a mathematical formulation, like the derivation of CAPM, it is interesting to review nevertheless.

Through the definition of these agents, they formulate a way to represent the demand for bitcoins for each one. Then, assuming the market is only composed of said agents, they develop an asset pricing model. Finally, regarding the conclusions, one of them was that due to the special characteristics of bitcoin investors in terms of risk aversion, the fact that economic variables appear to not explain a significant part of returns is not much of a surprise. As to the models themselves, they manage to estimate the impacts of different types of investors during low and high bitcoin price volatility regimes.

Lastly, the research done by (Bennett et al., 2023), although it does not fit into any of the two groups of studies proposed in this literature review, it provides an interesting view about different behavioral finance aspects that apply to decentralized finance[6]. They mention that asset pricing in rapidly evolving markets is better explained through behavioral finance, rather than through traditional finance theory. Factors like investor attention, sentiment, heuristics and biases, and network effects interact to form a highly volatile and dynamic market (Bennett et al., 2023). A par-

[6]Emerging financial technology, in which cryptocurrency could be considered.

ticularly compelling aspect about said research, is that presents a theoretical model of behavioral finance applications for asset pricing models related to decentralized finance, that could be taken into account when an initial proposal of factors is made.

2.1.3 Initial proposal of factors

Having reviewed the related bibliography, determinant factors will now be proposed, which could be part of the mathematical formulation for the derivation of future models. It is important to note that the factors mentioned correspond to a preliminary proposal, and the specific way of how they could be included in the formulation of the models will not be addressed in this section.

Despite the fact that (Grobys & Sapkota, 2019) do not find significant evidence as to relevant momentum payoffs, they evaluate only utilizing said factor as an investment strategy, but that does not mean that it does not explain the variability of cryptocurrency returns. So, in line with the conclusions outlined in the research of (Liu et al., 2022), which does find an importance on momentum, the first factors to take into consideration are the momentum related. In order to provide greater insight, said factors are in relation to past returns (i.e. past one week returns), however the temporal aspects of said elements should be evaluated, to determine which alternative leads to better results. This type of factors could help model the behavioral finance side of the cryptocurrency market.

Furthermore, following the research of (Liu et al., 2022), size related factors should also be studied. The inclusion of this type of variables in to a theoretic formulation could not be so straightforward, but it is important to take them into consideration because of their significance in explaining the returns of cryptocurrencies.

Other aspects that could be accounted for correspond to representing different type of investors. In this case, the types of investors could be selected according to different characteristics, like for instance, introducing different levels of risk aversion.

Despite the fact that behavioral finance applications could be seen as endless, in terms of the different factors that could be derived from this area. Following an approach similar to the last presented research (Bennett et al., 2023) in the literature review, it would be interesting to study the viability of incorporating some factors that are of behavioral nature. Some alternatives could be: investor sentiment, investor psychological biases, or movement of other assets like commodities or stocks.

Finally, though there might be a great variety of factors that could be added to the mathematical formulation, the aggregation of them does not ensure that the model derived from said problem will explain a significant portion of the variation of the returns of cryptocurrencies. That is the reason why it is important to study whether the inclusion of a factor, significantly enhances the explanatory capacity of the model.

Chapter 3

Objectives and Methodological Framework

3.1 Objectives

By distinguishing between general and specific objectives and ensuring they are concrete, verifiable, and free from methodological details, these objectives should effectively guide the dissertation research and provide clear benchmarks for evaluation upon completion.

3.1.1 General Objective

Create and confirm a Portfolio Markowitz-inspired model to evaluate cryptocurrency returns in order to improve comprehension and assist in making informed investment choices within the cryptocurrency sector.

3.1.2 Specific Objectives

1. Develop a theoretical framework based on Portfolio Markowitz theory, which includes mathematical equations and fundamental principles, to support the

creation of a model for analyzing cryptocurrency returns.

2. Gather and prepare necessary datasets on cryptocurrency returns, ensuring the data is accurate and appropriate for testing and validating the subject model.

3. Conduct an empirical investigation following the methodology outlined by Fama and French, 1993 to evaluate the predictive capacity and robustness of the proposed model in capturing the dynamics of cryptocurrency returns.

4. Develop statistical analyses, including the GRS hypothesis test and mean adjusted R-squared examination, accounting for the signs of mean regression coefficients, to evaluate the model's explanatory power and identify potential areas of improvement.

5. Combine results from model testing and statistical analysis to determine how well the model explains cryptocurrency return variations and its usefulness in shaping portfolio construction and risk management strategies.

6. Present suggestions for future research paths and practical applications for investors and financial institutions utilizing the knowledge obtained from the constructed model and empirical studies.

3.2 Methodological Framework

In order to solve the literature gap that was mentioned on chapter 1, a description of the methodological framework will be proposed in order to achieve the aforementioned general objective, and the ones established in section 3.1.

3.2.1 Data Retrieval

A crucial part of the investigation is the data that will be used to do the related tests in order to check the validity and overall performance of the model. Due to the nature of this investigation and the accessibility, the data provider that was be selected is *Yahoo Finance*.

Yahoo Finance has an integrated library in *python*, that can be used to retrieve a wide range of data, from prices to financial ratios related to specific companies, and other market data. It has data from about ten thousand cryptocurrencies, but the only problem is that the "symbols"[1] can not be retrieved directly from said library. In order to complete this task *Yahooquery* was used.

Yahooquery is a python interface to unofficial *Yahoo Finance* API endpoints. So in this case, the endpoint related to cryptocurrency symbols ordered by market capitalization was used to retrieve said cryptocurrencies. The maximum amount of cryptocurrencies that this interface allowed to retrieve was 250. The combination of these tools allowed to retrieve data of prices from 2014 onward of 250 cryptocurrencies ordered by intraday market capitalization[2].

3.2.2 Mathematical Formulation

Another part of the investigation that is essential is the derivation of the model. The general idea of the mathematical formulation comes from the derivation of the Capital Asset Pricing Model. Although the traditional optimization problem minimizes the variance of the portfolio, in this case an alternative approach that is also used will be taken.

Considering this scenario, the objective will be the utility function of a certain type of investor, which depends on the terminal wealth of said individual. The idea is to maximize this utility function subject to two constraints related to the initial and final wealth of the investor. The mathematical representation is as follows,

$$\max_{\boldsymbol{n}_j} E\left[U(w_j)\right] \ . \tag{3.1}$$

Subject to:

$$w_j = \boldsymbol{n}_j^{\intercal}\boldsymbol{x} + n_j^f \ , \tag{3.2}$$

$$\bar{w}_j = \boldsymbol{n}_j^{\intercal}\boldsymbol{P} + n_j^f P_f \ . \tag{3.3}$$

[1]Referring to the form cryptocurrencies are normally presented in exchanges, for example BTC-USD, which corresponds to the price of Bitcoin in US dollars.

[2]Market value of a cryptocurrency's stock at any given point during the trading day.

Where w_j and $\bar{w}_j$ are the terminal and initial wealth for investor type j, respectively. Then, $\boldsymbol{n}_j$ is the vector representing the amount investor type j purchases in each of N cryptocurrencies, and n_j^f the number of risk-free discount bonds with unit payoff purchased by investor type j. $\boldsymbol{P}$ is the vector of cryptocurrency prices, and P_f is the price of the discount bond.

Through the development of this theoretical formulation a formal model can be derived that explains the cross-section of returns of a certain cryptocurrency. But the latter is the general idea, the detail will be delved into in further sections of this dissertation.

3.2.3 Fama Mac-Beth Regressions

Shifting the focus to the empirical tests, the Fama Mac-Beth two-step regression is a commonly used technique in empirical finance for determining parameter estimates in asset pricing models. The technique calculates the betas and risk premiums for all risk factors believed to influence asset prices. The fundamental concept of the regression method is to predict the returns of assets by analyzing their factor exposures or characteristics that mirror exposure to a risk factor in each period.

To better understand the methodology, some equations need to be presented. The model formulated in this case states that the average excess return of a given cryptocurrency i is determined by the sensitivity of the cryptocurrency to the market risk factor β_{im} as well as the sensitivity to a "popular" factor β_{ip}.

$$\mu_i = \beta_{ip}\mu_p + \beta_{im}\mu_m \; . \tag{3.4}$$

The approach involves two consecutive stages. In the first step, time series estimates of the betas $(\hat{\beta}_{im}, \hat{\beta}_{ip})$ are calculated for the individual portfolios[3]. In the second step, these beta estimates are employed in a cross-sectional regression to obtain the estimates of the parameters of the regression $(\hat{\mu}_{p,t}, \hat{\mu}_{m,t})$, which are averaged over time, yielding the respective estimates $(\hat{\mu}_p, \hat{\mu}_m)$. These parameter estimates averages

[3]For the purpose of this dissertation portfolios are used, but the definition can also be with individual assets.

are finally compared statistically to their predicted values (Balvers, 2001).

One last detail that is important to mention, is that in each of the two steps the coefficient of determination is computed, which allows to know the percentage of the variability of data that is explained by the model in question.

3.2.4 Gibbons, Ross, and Shanken Test

Moreover, the statistical test outlined in (Gibbons et al., 1989) serves to assess the precision of asset pricing models. It is particularly employed to scrutinize whether the expected returns of a set of portfolios can be explained by their exposure to a common set of risk factors. Employing this test will facilitate the examination of the stated hypothesis, with a crucial emphasis on not rejecting the null hypothesis.

Following the explanation given in the paper itself, considering the following multivariate regression,

$$\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{\epsilon}_{it} \quad \forall i = 1, \ldots, N \ , \tag{3.5}$$

where $\tilde{r}_{it}$ is the excess return on asset i in period t; $\tilde{r}_{pt}$ is the excess return on the portfolio whose efficiency is being tested; and $\tilde{\epsilon}_{it}$ is the disturbance term for asset i on period t. The latter assuming that there is a given risk-free rate of interest R_{ft}, for each time period.

Then, if a particular portfolio is mean-variance efficient (i.e., it minimizes variance for a given level of expected return), the following first order condition must be satisfied for the given N assets:

$$E[\tilde{r}_{it}] = \beta_{ip}E[\tilde{r}_{pt}] \ . \tag{3.6}$$

Combining (3.5) with (3.6), yields the following parameter restriction, which can be stated in the form of a null hypothesis:

$$H_0 : \alpha_{ip} = 0, \quad \forall i = 1, \ldots, N \ .$$

This is the general explanation of the idea behind this test, and for simplicity, the detail of the equations related to the computation of the parameters that are used to carry out the test itself will be omitted.

3.2.5 Conclusions and recommendations

Ultimately, in 3.2.3 the results from the Fama-MacBeth Regressions and in 3.2.4 the GRS test will be examined to draw conclusions. This analysis aims to ascertain if the model effectively describes the cross-section of cryptocurrency returns and if the factors are statistically significant.

Despite the results, suggestions for further studies will be given. If the model proves to be accurate, its practical implications will be explored, offering valuable insights for both individual investors and financial institutions.

Chapter 4

Methodological Development

I need to establish: Type of study, Period and place where the research was done, Universe and sample, Methods, Variable selection, Assumptions, Procedures, Methods of data collection.

As it was already mentioned in previous sections of this dissertation, the investigation corresponds to an empirical finance study of a model that is derived from a theoretical formulation that was already presented in 3.2.2. The investigation was developed in the year 2024, in *Universidad de los Andes* from *Santiago, Chile*. In the following sections more detail will be presented as to every step showcased in the Methodological Framework relating the research itself.

4.1 Mathematical Formulation

Although the optimization problem was already presented in 3.2.2, it was the following.

$$\max_{\boldsymbol{n}_j} E\left[U(w_j)\right] \ . \tag{4.1}$$

Subject to:

$$w_j = \boldsymbol{n}_j^{\intercal}\boldsymbol{x} + n_j^f \ , \tag{4.2}$$

$$\bar{w}_j = \boldsymbol{n}_j^{\intercal}\boldsymbol{P} + n_j^f P_f \ . \tag{4.3}$$

Where in summary the variables represent:

- w_j and $\bar{w}_j$ are the terminal and initial wealth for investor type j.
- $\boldsymbol{n}_j$ is the vector representing the amount investor type j purchases in each of N cryptocurrencies.
- n_j^f the number of risk-free discount bonds with unit payoff purchased by investor type j.
- $\boldsymbol{P}$ is the vector of cryptocurrency prices, and P_f is the price of the discount bond.

An important aspect is that this model derivation is done originally in (Luo & Balvers, 2017). The model formulated in this dissertation is based on said paper, but in this case it is applied to cryptocurrencies. Also, in the paper the detail of the math is given, but there are some steps that are not shown in a level of detail that allows for a full understanding of the process, so in those cases mathematical intuition was required to obtain the required results in order to derive the model. Said steps will be mentioned in this section.

4.1.1 Math Detail

For the model there are two types of investors: the unrestricted investors (U), and the restricted investors (R). In this case, the restricted investors invest solely in cryptocurrencies that hold a dominant position in terms of popularity and market capitalization.

In the traditional Capital Asset Pricing Model, the unrestricted investor fully consumes terminal wealth, with w_U being terminal wealth of the unrestricted investor. For the said investor, the problem is as follows,

$$\max_{\mathbf{n_U}} \mathrm{E}\left[U(w_U)\right] \ . \tag{4.4}$$

Subject to:

$$w_U = \boldsymbol{n}_U^\intercal \boldsymbol{x} + n_U^f \ , \tag{4.5}$$

$$\bar{w}_U = \boldsymbol{n}_U^\intercal \boldsymbol{P} + n_U^f P_f \ . \tag{4.6}$$

The following step was not detailed in the paper. From (4.6), the following conclusion can be drawn,

$$n_U^f = \frac{1}{P_f} \left(\bar{w}_U - \boldsymbol{n}_U^\intercal \boldsymbol{P} \right) \ .$$

Then, substituting the expression in (4.5),

$$w_U = \boldsymbol{n}_U^\intercal \boldsymbol{x} + \bar{w}_U \frac{1}{P_f} - \boldsymbol{n}_U^\intercal \underbrace{\frac{\boldsymbol{P}}{P_f}}_{\boldsymbol{p}} = \frac{\bar{w}_U}{P_f} + \boldsymbol{n}_U^\intercal (\boldsymbol{x} - \boldsymbol{p}) \tag{4.7}$$

Substituting (4.7) in (4.4), and computing the derivative.

$$\frac{d\mathrm{E}}{dw_U} = \mathrm{E} \left[U'(w_U)(\boldsymbol{x} - \boldsymbol{p}) \right] = 0 \ .$$

Which corresponds to the first order condition. Now, taking into account that $\boldsymbol{x} \sim \mathcal{N}(\bar{\boldsymbol{x}}, \boldsymbol{\Sigma})$, which is that the payoff vector is multivariate normally distributed, applying the definition of covariance (see appendix A.1) the following expression can be deduced, which with the lemma application, were not detailed in the paper,

$$\begin{aligned}
\mathrm{E} \left[U'(w_U)(\boldsymbol{x} - \boldsymbol{p}) \right] &= E \left[(U'(w_U) - E \left[U'(w_U) \right])(\boldsymbol{x} - \boldsymbol{p} - E \left[\boldsymbol{x} - \boldsymbol{p} \right]) \right] + E \left[U'(w_U) \right] E \left[\boldsymbol{x} - \boldsymbol{p} \right] \ , \\
&= E \left[(U'(w_U) - E[U'(w_U)])(\boldsymbol{x} - \bar{\boldsymbol{x}}) \right] + E \left[U'(w_U) \right] (\bar{\boldsymbol{x}} - \boldsymbol{p}) \ , \\
&= E \left[U'(w_U)(\boldsymbol{x} - \bar{\boldsymbol{x}}) - E[U'(w_U)](\boldsymbol{x} - \bar{\boldsymbol{x}}) \right] + E \left[U'(w_U) \right] (\bar{\boldsymbol{x}} - \boldsymbol{p}) \ , \\
&= E \left[U'(w_U)(\boldsymbol{x} - \bar{\boldsymbol{x}}) \right] - E[U'(w_U)] E \left[\boldsymbol{x} - \bar{\boldsymbol{x}} \right] + E \left[U'(w_U) \right] (\bar{\boldsymbol{x}} - \boldsymbol{p}) \ , \\
\mathrm{E} \left[U'(w_U)(\boldsymbol{x} - \boldsymbol{p}) \right] &= E \left[U'(w_U)(\boldsymbol{x} - \bar{\boldsymbol{x}}) \right] + E \left[U'(w_U) \right] (\bar{\boldsymbol{x}} - \boldsymbol{p}) = 0 \ .
\end{aligned}$$

Then the following equality can be defined,

$$-E \left[U'(w_U)(\boldsymbol{x} - \bar{\boldsymbol{x}}) \right] = E \left[U'(w_U) \right] (\bar{\boldsymbol{x}} - \boldsymbol{p})$$

Applying the lemma in appendix A.2,

$$\begin{aligned} -E\left[U''(w_U)\right]\boldsymbol{\Sigma}\boldsymbol{n}_U &= E\left[U'(w_U)\right](\bar{\boldsymbol{x}}-\boldsymbol{p}) \ , \\ \bar{\boldsymbol{x}}-\boldsymbol{p} &= \frac{-E\left[U''(w_U)\right]}{E\left[U'(w_U)\right]}\boldsymbol{\Sigma}\boldsymbol{n}_U \ , \\ \bar{\boldsymbol{x}}-\boldsymbol{p} &= \theta_U\boldsymbol{\Sigma}\boldsymbol{n}_U \ . \end{aligned} \tag{4.8}$$

Where $\theta_U = -E\left[U''(w_U)\right]/E\left[U'(w_U)\right]$ is analogous to absolute risk aversion[1], which depends on the initial wealth of investor U and other model. $\boldsymbol{\Sigma}$ is the covariance matrix for risky asset payoffs and $\bar{\boldsymbol{x}}$ the expected payoffs of risky assets.

For investor type R the problem is of similar nature,

$$\max_{\mathbf{n_R}} \mathrm{E}\left[U(w_R)\right] \ . \tag{4.9}$$

Subject to:

$$w_R = \boldsymbol{n}_R^{\intercal}\boldsymbol{x} + n_R^f \ , \tag{4.10}$$

$$\bar{w}_R = \boldsymbol{n}_R^{\intercal}\boldsymbol{P} + n_R^f P_f \ . \tag{4.11}$$

Where $\boldsymbol{n}_R$ is the vector of the shares of cryptocurrencies that investor R purchases that comply with their preferences. Then, following the same procedure as before.

$$\theta_R\boldsymbol{\Sigma}_P\boldsymbol{n}_R = \bar{\boldsymbol{x}}_P - \boldsymbol{p}_P \ . \tag{4.12}$$

Where the matrix of asset payoff covariances is partitioned into popular (P) and non-popular (N) cryptocurrencies.

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_P & \boldsymbol{\Sigma}_{PN} \\ \boldsymbol{\Sigma}_{NP} & \boldsymbol{\Sigma}_N \end{bmatrix} \tag{4.13}$$

Where $\boldsymbol{\Sigma}_N$ represents the payoff covariance of all cryptocurrencies that are "non-popular" or have small market capitalization, and $\bar{\boldsymbol{x}}_N$ and $\boldsymbol{p}_N$ are the vectors of mean payoffs and prices, respectively, of the "non-popular" cryptocurrencies.

[1]Tendency of individuals to prefer outcomes with low uncertainty over those with high uncertainty, even if the average outcomes of the latter is equal to or higher in monetary value that the more certain outcome (Pratt, 1964).

Assuming q_U investors of type U and q_R investors of type R, the demand for cryptocurrencies may be obtained and set equal to the exogenous supply of cryptocurrencies $\bar{\boldsymbol{n}} = (\bar{\boldsymbol{n}}_N, \bar{\boldsymbol{n}}_P)^\intercal$, and to zero for the risk-free asset, yielding the conditions for market equilibrium.

$$\bar{\boldsymbol{n}} = q_U \boldsymbol{n}_U + q_R \boldsymbol{n}_R, \quad 0 = q_U n_U^f + q_R n_R^f \,. \tag{4.14}$$

Reorganizing equations (4.12) and (4.8) yields the following,

$$\boldsymbol{n}_U = (\theta_U \boldsymbol{\Sigma})^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{p}), \quad \boldsymbol{n}_R = (\theta_R \boldsymbol{\Sigma}_P)^{-1} (\bar{\boldsymbol{x}}_P - \boldsymbol{p}_P) \,.$$

The following step was not detailed. Note that $\boldsymbol{n}_R$ can be represented in the following form,

$$\boldsymbol{n}_R = \theta_R^{-1} \begin{bmatrix} \boldsymbol{\Sigma}_P^{-1} & 0 \\ 0 & 0 \end{bmatrix} (\bar{\boldsymbol{x}} - \boldsymbol{p}) = \begin{bmatrix} \boldsymbol{I} \\ 0 \end{bmatrix} (\boldsymbol{\Sigma}_P \theta_R)^{-1} \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} (\bar{\boldsymbol{x}} - \boldsymbol{p}) \,.$$

Substituting in (4.14) yields the following,

$$\bar{\boldsymbol{n}} = \left((\boldsymbol{\Sigma} \theta_U / q_U)^{-1} + \begin{bmatrix} \boldsymbol{I} \\ 0 \end{bmatrix} (\boldsymbol{\Sigma}_P \theta_R / q_R)^{-1} \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} \right) (\bar{\boldsymbol{x}} - \boldsymbol{p}) \,. \tag{4.15}$$

From where we want to isolate the expression $\bar{\boldsymbol{x}} - \boldsymbol{p}$, then is necessary to compute the inverse of the expression in parenthesis. The latter can be done using an identity (Söderström, 2002) that says the following. Given matrices $\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3$ y $\boldsymbol{X}_4$, with $\boldsymbol{X}_1$, $\boldsymbol{X}_4$ having an inverse, the following equality is satisfied.

$$\left(\boldsymbol{X}_1^{-1} + \boldsymbol{X}_2 \boldsymbol{X}_4^{-1} \boldsymbol{X}_3 \right)^{-1} = \boldsymbol{X}_1 + \boldsymbol{X}_1 \boldsymbol{X}_2 \left(\boldsymbol{X}_4 + \boldsymbol{X}_3 \boldsymbol{X}_1 \boldsymbol{X}_2 \right)^{-1} \boldsymbol{X}_3 \boldsymbol{X}_1 \,. \tag{4.16}$$

This step and the subsequent one were not detailed. Substituting the terms in (4.16), yields the following,

$$
\begin{aligned}
&\left((\boldsymbol{\Sigma}\theta_U/q_U)^{-1} + \begin{bmatrix} \boldsymbol{I} \\ 0 \end{bmatrix} (\boldsymbol{\Sigma}_P\theta_R/q_R)^{-1} \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} \right)^{-1} \\
&= \boldsymbol{\Sigma}\theta_U/q_U - \boldsymbol{\Sigma}\theta_U/q_U \begin{bmatrix} \boldsymbol{I} \\ 0 \end{bmatrix} \left(\boldsymbol{\Sigma}_P\theta_R/q_R + \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} \boldsymbol{\Sigma}\theta_U/q_U \begin{bmatrix} \boldsymbol{I} \\ 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} \boldsymbol{\Sigma}\theta_U/q_U \ , \\
&= \boldsymbol{\Sigma}\theta_U/q_U - \boldsymbol{\Sigma}\theta_U/q_U \begin{bmatrix} \boldsymbol{I} \\ 0 \end{bmatrix} (\boldsymbol{\Sigma}_P\theta_R/q_R + \boldsymbol{\Sigma}_P\theta_U/q_U)^{-1} \begin{bmatrix} \boldsymbol{I} & 0 \end{bmatrix} \boldsymbol{\Sigma}\theta_U/q_U \ , \\
&= \theta_U/q_U \left(\boldsymbol{\Sigma} - \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{\Sigma}_P^{-1} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\Sigma} \right) \ .
\end{aligned}
$$

Then, substituting the expression in (4.15) yields the following,

$$
\begin{aligned}
(\bar{\boldsymbol{x}} - \boldsymbol{p}) &= \theta_U/q_U \left(\boldsymbol{\Sigma} - \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{\Sigma}_P^{-1} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{\Sigma} \right) \bar{\boldsymbol{n}} \ , \\
&= \theta_U/q_U \left(\boldsymbol{\Sigma} - \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{PN} \\ 0 & 0 \end{bmatrix} \right) \bar{\boldsymbol{n}} \ , \\
&= \theta_U/q_U \left(\boldsymbol{\Sigma}\bar{\boldsymbol{n}} - \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma} \begin{bmatrix} \bar{\boldsymbol{n}}_N + \boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{PN}\bar{\boldsymbol{n}}_P \\ 0 \end{bmatrix} \right) \ , \\
&= \theta_U/q_U \left(\boldsymbol{\Sigma}\bar{\boldsymbol{n}} - \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma}\bar{\boldsymbol{n}} + \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma} \begin{bmatrix} -\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{PN}\bar{\boldsymbol{n}}_P \\ \bar{\boldsymbol{n}}_P \end{bmatrix} \right) \ , \\
&= \theta_U/q_U \left(\frac{\theta_R/q_R}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma}\bar{\boldsymbol{n}} + \frac{\theta_U/q_U}{\theta_U/q_U + \theta_R/q_R} \boldsymbol{\Sigma} \begin{bmatrix} -\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{PN}\bar{\boldsymbol{n}}_P \\ \bar{\boldsymbol{n}}_P \end{bmatrix} \right) \ , \\
&= \left(\frac{1}{q_U/\theta_U + q_R/\theta_R} \boldsymbol{\Sigma}\bar{\boldsymbol{n}} + \frac{1}{q_U/\theta_U + q_R/\theta_R} \frac{q_R/\theta_R}{q_U/\theta_U} \boldsymbol{\Sigma} \begin{bmatrix} -\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{PN}\bar{\boldsymbol{n}}_P \\ \bar{\boldsymbol{n}}_P \end{bmatrix} \right) \ , \\
&= \frac{1}{q_U\bar{w}_U/\rho_U + q_R\bar{w}_R/\rho_R} \boldsymbol{\Sigma}\bar{\boldsymbol{n}} + \frac{1}{q_U\bar{w}_U/\rho_U + q_R\bar{w}_R/\rho_R} \frac{q_R\bar{w}_R/\rho_R}{q_U\bar{w}_U/\rho_U} \boldsymbol{\Sigma} \begin{bmatrix} -\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_{PN}\bar{\boldsymbol{n}}_P \\ \bar{\boldsymbol{n}}_P \end{bmatrix} \ , \\
&= \gamma\boldsymbol{\Sigma}\bar{\boldsymbol{n}} + \delta\boldsymbol{\Sigma}\bar{\boldsymbol{n}}_K \ .
\end{aligned}
\tag{4.17}
$$

Where $\bar{\boldsymbol{n}}_K$ represents the known cryptocurrency portfolio. Now, (4.17) must be converted into an expression for expected returns rather than expected net payoffs.

Given that $P_f = 1/(1+r_f)$, the following can be defined,

$$(1+r_i^s) = \frac{x_i}{P_i} \Leftrightarrow x_i - \frac{P_i}{P_f} = P_i(1+r_i^s) - P_i(1+r_f) = P_i(r_i^s - r_f) \; .$$

Then, defining the excess return as $r_i = r_i^s - r_f$, and given that in Equation (4.17) the expression to the left of the equality is represented as an average, it follows that $\mu_i = \mu_i^s - r_f$. In addition, since $1+r_i^s = x_i/P_i$, the covariance matrix for the payoffs of the cryptocurrencies $\mathbf{\Sigma}$ can be represented in terms of the returns as $\sigma_{ij} = \Sigma_{ij}/P_iP_j$. Thus, for a specific element of Equation (4.17), it can be stated that,

$$\begin{aligned} P_i\mu_i &= \gamma\Sigma_{im} + \delta\Sigma_{ip} \\ \mu_i &= \gamma P_m\sigma_{im} + \delta P_p\sigma_{ip} \end{aligned} \tag{4.18}$$

Where m represents the market, $P_m = q_m\bar{w}_M = q_U\bar{w}_U + q_R\bar{w}_R$ is the cost of the market portfolio, and P_p is the cost of the popular portfolio. Now, given (4.18), μ_m and μ_p can be defined, which correspond to the mean returns of the market and popular portfolios, respectively.

$$\mu_m = \gamma P_m\sigma_m^2 + \delta P_p\sigma_{mp} \quad ; \quad \mu_p = \gamma P_m\sigma_{mp} + \delta P_p\sigma_p^2 \; .$$

Solving the system of equations for γP_m y δP_p yields the following,

$$\delta P_p = \frac{\sigma_{mp}\mu_m - \sigma_m^2\mu_p}{\sigma_{mp}^2 - \sigma_p^2\sigma_m^2} \quad ; \quad \gamma P_m = \frac{\sigma_{mp}\mu_p - \sigma_p^2\mu_m}{\sigma_{mp}^2 - \sigma_p^2\sigma_m^2} \; .$$

This step was not detailed. Substituting in (4.18) yields,

$$\begin{aligned} \mu_i &= \frac{\sigma_m^2\sigma_{ip} - \sigma_{mp}\sigma_{im}}{\sigma_p^2\sigma_m^2 - \sigma_{mp}^2}\mu_p + \frac{\sigma_p^2\sigma_{im} - \sigma_{mp}\sigma_{ip}}{\sigma_p^2\sigma_m^2 - \sigma_{mp}^2}\mu_m \; , \\ \mu_i &= \beta_{ip}\mu_p + \beta_{im}\mu_m \; . \end{aligned} \tag{4.19}$$

Where β_{ib} and β_{ip} are the population values of the slope estimates for a linear regression of the return of asset i on the market portfolio return and the popular portfolio return.

Some comments

An attempt was made to modify the original derivation process that uses two type of investor, by adding a third one. But the issue with this is that the step in which (Söderström, 2002) identity is used, there are problems respecting the amount of matrices needed to use said identity. Assuming there are three types of investors U, R_1, and R_2, the optimization problems do not change, but the covariance matrix does,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{P_1} & \mathbf{\Sigma}_{P_1 P_2} & \mathbf{\Sigma}_{P_1 N} \\ \mathbf{\Sigma}_{P_2 P_1} & \mathbf{\Sigma}_{P_2} & \mathbf{\Sigma}_{P_2 N} \\ \mathbf{\Sigma}_{N P_1} & \mathbf{\Sigma}_{N P_2} & \mathbf{\Sigma}_{N} \end{bmatrix} .$$

Then, the terms respecting the exogenous supply of cryptocurrencies would be as follows,

$$\bar{\boldsymbol{n}} = q_U \boldsymbol{n}_U + q_{R_1} \boldsymbol{n}_{R_1} + q_{R_2} \boldsymbol{n}_{R_2}, \quad 0 = q_U n_U^f + q_{R_1} n_{R_1}^f + q_{R_2} n_{R_2}^f . \tag{4.20}$$

Defining the following,

$$\boldsymbol{n}_U = (\theta_U \mathbf{\Sigma})^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{p}), \quad \boldsymbol{n}_{R_1} = (\theta_{R_1} \mathbf{\Sigma}_{P_1})^{-1} (\bar{\boldsymbol{x}}_{P_1} - \boldsymbol{p}_{P_1}), \quad \boldsymbol{n}_{R_2} = (\theta_{R_2} \mathbf{\Sigma}_{P_2})^{-1} (\bar{\boldsymbol{x}}_{P_2} - \boldsymbol{p}_{P_2}) .$$

This would imply that if those therms are replaced in (4.20), three new matrices appear in the expression in which the identity presented in (Söderström, 2002) is applied. Making it impossible to apply said identity, cause it requires only four matrices, not six.

4.2 Data

As it was already mentioned on 3.2.1, the source of the data corresponds to the *Yahoo Finance* library in *Python*, that is used in conjunction with *Yahooquery*, where the latter is utilized for the retrieval of cryptocurrency symbols.

Regarding the universe and the sample, the initial set comprised 250 cryptocurrencies. Stablecoins were excluded from this set, resulting in a total of 226 cryptocurrencies paired with the US Dollar. An attempt was made to eliminate stablecoins using a criterion based on the standard deviation and mean of their returns. However, for accuracy, the identification and removal of these stablecoins were ultimately done manually. Conversely the sample, like it was already mentioned in 3.2.1, were 250 cryptocurrencies ordered by intraday market capitalization. The dates ranges of the data are from September, 2014 till March, 2024.

4.2.1 Market Index

A crypto market index was utilized in order to have a the market factor in the model. Said factor was the *Crypto200 ex BTC Index by Solactive*, that is comprised by a volume weighted average price[2] on the top 200 cryptocurrencies except for Bitcoin and stablecoins, by market capitalization, in USD.

Add source.

One important detail is that the data for this index has been available only from 2019 onward. This limitation affected the sample size, necessitating a reduction to conduct the respective regressions.

4.2.2 Return computation

All the returns in this investigation correspond to weekly returns. To compute them, the market index data was retrieved first, as it had fewer available dates than the cryptocurrency data. Weekly values of the market index were collected, and the returns were subsequently calculated.

Having the dates of the market returns, for the calculations of the cryptocurrency returns, daily data was used, and the dates were filtered in order for them to match the ones of the market data. Then, with the filtered dates, the weekly returns were computed for each cryptocurrency in the dataset.

[2] Trading metric that calculates the average price of an asset based on both the trading volume and the price. It gives more weight to trades with higher volume. It provides a more accurate representation of the average price considering the trading activity.

4.2.3 "Popular" Factor Estimation

In order to build the "popular" factor that is derived in the model, a portfolio comprised of these type of cryptocurrencies was built (see table A.1 for the detail). In the other hand, the "not popular" portfolio was built using the rest of the cryptocurrencies that were not in the "popular" portfolio. Then, the returns for each portfolio were calculated using a value weighted fashion, with their respective market capitalization's.

One aspect that is important to detail that also was used in further stages of the data handling, was that the dates in which the cryptocurrencies have data vary depending on each one. The latter because there are cryptocurrencies that are newer than others. This affects directly the computation of portfolio returns because there will be dates were not all the cryptocurrencies in the portfolio will have available returns.

So, for the portfolio returns, the weights are computed for each date based on the availability of returns for the cryptocurrencies in the portfolio on that date. This method was used for all the portfolio computations on this investigation.

After calculating the returns for each portfolio, statistical analysis was conducted. First, a histogram of the returns was created to observe whether the distribution resembled a normal distribution. Analyzing figure 4.1, the resemblance is quite similar to a normal distribution.

Then, t-student hypothesis test was made, in order to check if there is a significant difference in the returns of both portfolios, because in otherwise, the idea of a "popular" factor would loose credibility in explaining the cross section of returns.

The t-statistic in a t-test indicates how many standard errors the sample mean is from the sample mean of another group. The sign of the t-statistic can also indicate the direction of the difference, if one exists. In this case, a positive sign suggests that the mean returns of the "not popular" portfolio are greater than those of the "popular" portfolio, while a negative sign suggests the opposite. One important detail is that for this test the whole date range was used that is available in the cryptocurrency dataset, which is large that the date range of the market index data.

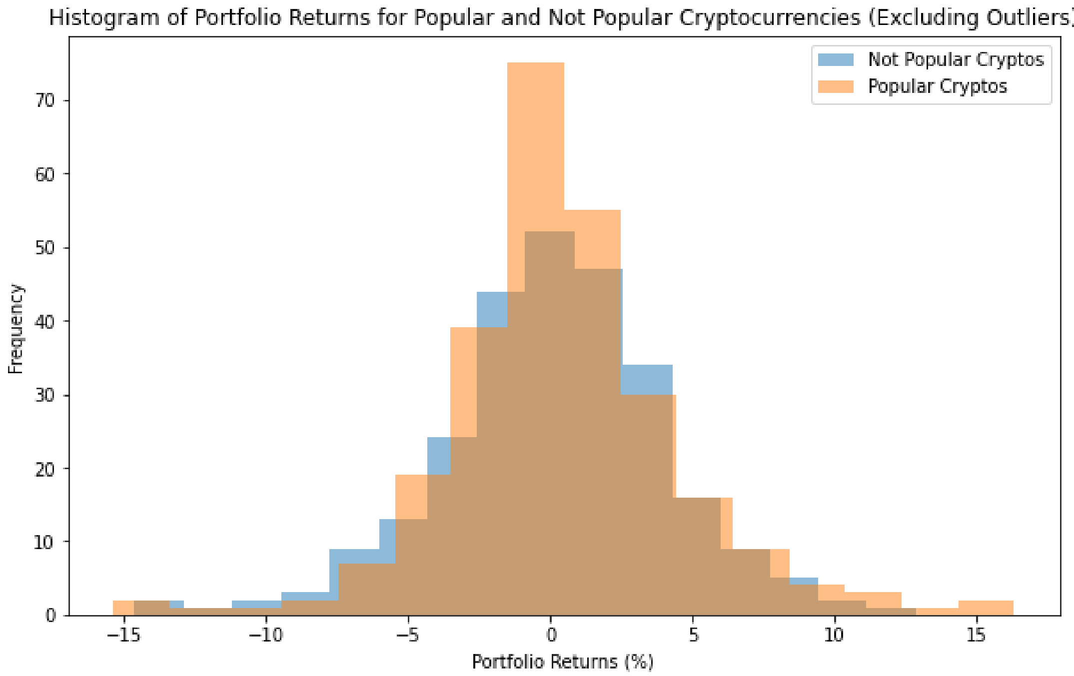

Figure 4.1: Histogram of portfolio returns for "popular" and "not popular" cryptocurrencies.

Table 4.1: T-Test Results.

Statistic	P-value
1.71	0.088

The null hypothesis in this case is that there is no significant difference in portfolio returns, while the alternative hypothesis suggests otherwise. From table 4.1, it can be concluded that there is sufficient statistical evidence to reject the null hypothesis in favor of the alternative. Specifically, the results indicate that the mean portfolio return of the "not popular" portfolio is greater than that of the "popular" portfolio.

The last "test" that was done corresponded to the computation of the mean portfolio returns of the whole sample, using all the dates available in the cryptocurrency data. From table 4.2, it can be seen that there is an important difference in mean

Table 4.2: Mean Portfolio Returns.

Portfolio	Mean Return (%)
Not Popular	10.23%
Popular	1.54%

returns, when taking into consideration the complete date range.

With all the necessary statistical tests completed, the process of constructing the "popular" factor could commence. Initially, a zero investment portfolio was constructed utilizing the difference in returns between the"popular" and "not popular" portfolios. Subsequently, a regression was performed, with the market returns serving as the independent variable and the zero investment portfolio returns as the dependent variable. Ultimately, the estimation of the "popular" factor corresponds to the residual error of this regression model.

4.3 Portfolio Building

In (Fama & French, 2004) it is explained that while CAPM can be a useful tool, it often performs better when applied to portfolios rather than individual assets, mainly because of the diversification of idiosyncratic risk. For that reason, portfolios of cryptocurrencies were built in order to test the validity of the model.

Tu build said portfolios, regressions were estimated for every cryptocurrency, using the "popular" factor as the independent variable, and the returns of an individual cryptocurrency as the dependent variable. So, for every cryptocurrency, a beta was estimated for the "popular" factor.

Using those estimated betas, all the cryptos were arranged from higher to lower value. Then, the portfolios were formed choosing from said list in descending order, based on a certain number of cryptocurrencies for each one.

A really important detail is the one mentioned in section 4.2.3 that is related to the forming of portfolios and the availability of returns on certain dates. In this part, that phenomena impacted directly on the number of cryptos that every portfolio has, because it had to be in a way that all the portfolios could have returns in all the range of dates so that regressions could be done later on. A lot of alternatives were tested, but the number that allowed for the regressions was a total of 26 cryptocurrencies in each portfolio.

Although said quantity meant that there was one portfolio with less cryptos than others, the latter did not poise a problem to carry on with the regressions. In detail portfolios 1 through 8 had 26 cryptocurrencies, and the ninth portfolio had 18.

Finally, for each portfolio, the respective returns were computed using value weighted and equally weighted methods.

4.4 Fama Mac-Beth Regressions

In section 3.2.3 a general overview of the methodology was presented, but in practice, another approach was taken that is similar to the one already mentioned. In this case, the first pass is the GRS test presented on section 3.2.4, where a p-value is obtained to determine if the null hypothesis is rejected or not, it is important to remember that the objective is not to reject said hypothesis.

The second pass consisted of two consecutive stages, as detailed in section 3.2.3. First, one regression is estimated for the average returns, were an adjusted R-squared is obtained as well as the factor loadings for every factor with its respective t statistic value. Second, a cross-sectional regression is estimated for every week, from were an average R-squared is obtained with the estimations of the factors loadings and their respective t statistics. From this step, coefficient estimates are obtained for every factor and the intercept, and also an average R-squared, which is normally the one that is used for empirical studies.

The ideal outcome in this section is to achieve a high adjusted R-squared, with an intercept that is not statistically significant, and statistically significant factor coefficients.

Chapter 5

Results

For the testing of the model the methods presented in previous chapters were utilized, and the results obtained will be reviewed. The Fama Mac-Beth methodology was made on various time periods to test if the model presented different results based on that factor. The time windows were selected based on if the crypto market was on a bull state on said weeks.

For all the dataset the results obtained were the following.

References

Balvers, R. J. (2001). *Foundations of asset pricing*. West Virginia University. https://balvers.azurewebsites.net/notes.html

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

Bennett, D., Mekelburg, E., & Williams, T. (2023). Befi meets defi: A behavioral finance approach to decentralized finance asset pricing. *Research in International Business and Finance*, *65*. https://doi.org/10.1016/j.ribaf.2023.101939

Cai, C. X., & Zhao, R. (2024). Salience theory and cryptocurrency returns. *Journal of Banking & Finance*, *159*, 107052.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57–82.

Cong, L. W., Li, Y., & Wang, N. (2020). Tokenomics: Dynamic Adoption and Valuation. *The Review of Financial Studies*, *34*(3), 1105–1155. https://doi.org/10.1093/rfs/hhaa089

Cosemans, M., & Frehen, R. (2021). Salience theory and stock prices: Empirical evidence. *Journal of Financial Economics*, *140*(2), 460–483. https://doi.org/https://doi.org/10.1016/j.jfineco.2020.12.012

Diaconaşu, D.-E., Mehdian, S., & Stoica, O. (2022). An analysis of investors' behavior in bitcoin market. *PLOS ONE*, *17*(3), 1–18. https://doi.org/10.1371/journal.pone.0264522

Erfanian, S., Zhou, Y., Razzaq, A., Abbas, A., Safeer, A. A., & Li, T. (2022). Predicting bitcoin (btc) price in the context of economic theories: A machine learning approach. *Entropy*, *24*(10). https://doi.org/10.3390/e24101487

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.

Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, *18*(3), 25–46.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

George G., R. (2003). *An introduction to probability and statistical inference.* Academic Press. https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=195530&lang=es&site=ehost-live

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, *57*(5), 1121–1152. Retrieved May 13, 2024, from http://www.jstor.org/stable/1913625

Gregoriou, A. (2019). Cryptocurrencies and asset pricing. *Applied Economics Letters*, *26*(12), 995–998.

Grobys, K., & Sapkota, N. (2019). Cryptocurrencies and momentum. *Economics Letters*, *180*, 6–10. https://doi.org/https://doi.org/10.1016/j.econlet.2019.03.028

Hasan, M., Naeem, M. A., Arif, M., Shahzad, S. J. H., & Vo, X. V. (2022). Liquidity connectedness in cryptocurrency market. *Financial Innovation*, *8*. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-021-00308-3

Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics*, *34*(7), 1308–1321. https://doi.org/https://doi.org/10.1016/j.tele.2016.05.005

Hull, J. (2012). *Options, futures , and other derivatives*. Prentice Hall.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Koutmos, D., & Payne, J. E. (2021). Intertemporal asset pricing with bitcoin. *Review of Quantitative Finance and Accounting*, *56*, 619–645.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *Journal of Finance*, *20*(4), 587–615.

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

Long, H., Zaremba, A., Demir, E., Szczygielski, J. J., & Vasenin, M. (2020). Seasonality in the cross-section of cryptocurrency returns. *Finance Research Letters*, *35*. https://doi.org/10.1016/j.frl.2020.101566

Luo, H. A., & Balvers, R. J. (2017). Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, *52*(1), 365–399.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, 768–783.

Prat, J., Danos, V., & Marcassa, S. (2019). Fundamental pricing of utility tokens. https://api.semanticscholar.org/CorpusID:208192945

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1-2), 122–136.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, *19*(3), 425–442.

Shen, D., Urquhart, A., & Wang, P. (2020). A three-factor pricing model for cryptocurrencies. *Finance Research Letters*, *34*. https://doi.org/10.1016/j.frl.2019.07.021

Sockin, M., & Xiong, W. (2020). A model of cryptocurrencies. (26816). https://doi.org/10.3386/w26816

Söderström, T. (2002). *Discrete-time stochastic systems: Estimation and control*. Springer.

Song, G. (2023). Size premium or size discount? – a dynamic capital mobility based interpretation. *Studies in Economics and Finance*, *40*(2), 266–285. https://www.emerald.com/insight/content/doi/10.1108/SEF-04-2022-0211/full/html

Thoma, A. (2020). A prospect theory model for predicting cryptocurrency returns. *Available at SSRN 3753530*.

Appendix A

Technical details, tables, and others

A.1 Covariance Definition

The definition used in the derivation of the model is the one that is presented in (George G., 2003) that states the following, assuming X and Y are random variables that have finite expectations,

$$Cov(X,Y) = E\left[(X - E(X))(Y - E(Y)\right] = E\left[XY\right] - (E\left[X\right])(E\left[Y\right])\ . \tag{A.1}$$

A.2 Stein's Lemma

A.3 Popular Portfolio

Name of Cryptocurrency	**Symbol**
Bitcoin	BTC-USD
Ethereum	ETH-USD
Binance Coin	BNB-USD
Solana	SOL-USD
Ripple	XRP-USD
Cardano	ADA-USD
Polkadot	DOT-USD
Chainlink	LINK-USD
Polygon	MATIC-USD

Table A.1: List of Popular Cryptocurrencies and their Symbols